The Dissertation Committee for Jacob Wesley Malcom Certifies that this is the approved version of the following dissertation:

FUNCTIONAL GENOMICS OF A MODEL ECOLOGICAL SPECIES, *DAPHNIA PULEX*

Committee:

Mathew Leibold, Supervisor

Thomas Juenger, Co-Supervisor

Hans Hoffman

Mikhail Matz

Michael Pfrender

FUNCTIONAL GENOMICS OF A MODEL ECOLOGICAL SPECIES, *DAPHNIA PULEX*

by

Jacob Wesley Malcom, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2013

Dedication

This work is dedicated to my loving wife, Anna, without whom I would be infinitely less happy.

Acknowledgements

I thank my advisors, Mathew Leibold and Tom Juenger, for their advice, patience, and support, and my dissertation committee for their feedback and support. The work underlying this dissertation was funded by grants from the University of Texas at Austin Ecology, Evolution, and Behavior program, a faculty grant to M. Leibold, and an NSF DEB grant to M. Leibold. In addition to the research encompassed by this dissertation, additional (but ancillary) research was facilitated by undergraduate assistants, including Kelly Likos, Skylar Netherland, and Nathan Wisnoski; many thanks to them for their valiant efforts. Last, many thanks to various individuals for providing useful feedback that improved the quality of the dissertation.

FUNCTIONAL GENOMICS OF A MODEL ECOLOGICAL SPECIES, *DAPHNIA PULEX*

Jacob Wesley Malcom, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Mathew A. Leibold

Co-supervisor: Thomas E. Juenger

Determining the molecular basis of heritable variation in complex, quantitative ecologically important traits will provide insight into the proximate mechanisms driving phenotypic and ecological variation, and the molecular evolutionary history of these traits. Furthermore, if the study organism is a "keystone species" whose presence or absence shapes ecological communities, then we extend our understanding of the effects of molecular variation to the level of communities. I examined the molecular basis of variation in 32 ecologically important traits in the freshwater pond keystone species *Daphnia pulex*, and identified thousands of candidate genes for which variation may affect not just *Daphnia* phenotypes, but the structure of communities. I extended the basic results to address two questions: what genes are associated with the offspring size-number trade-off in *Daphnia*; and can we identify candidate "keystone gene networks" for which variation may have a particularly strong influence on eco-evolutionary dynamics of limnetic communities? I found that different genes, with different biological functions, are associated with the trade-off in subsequent broods, and propose a model linking evolutionary frameworks to molecular biological functions. Next I found that

quantitative genetic variation in keystone traits appears to co-vary with the selection regimes to which *Daphnia* is subject, and identified two candidate gene networks that may underpin this genetic variation. Not only do these results provide a host of molecular hypotheses to be tested as *Daphnia* matures as a model genomic organism, but they also suggest models that link molecular research with broader themes in ecology, evolution, and behavior.

Table of Contents

List of Tables

List of Figures

Introduction

Explaining the origin and maintenance of biological diversity is the long-standing goal of evolution, ecology, and behavior research. This broad problem spans multiple levels of biological organization, from genetic sequence to the biodiversity of the entire biosphere; our goal is to understand how variation at one level percolates to affect variation at either higher or lower levels of organization. How does genetic variation affect phenotypic variation, and by extension, how does this phenotypic variation affect ecological interactions? Conversely, how do evolutionary processes such as mutation, selection, recombination, and drift affect the alleles present in subsequent generations, which shape possible phenotypic distributions and therefor, the range of ecological interactions?

The problem of connections across multiple levels of organization is succinctly captured by Dawkins's "extended phenotype", the idea that the effects of genetic variants extend well beyond the individual carrying the gene (Dawkins 1982; see also *niche construction,* Odling-Smee et al. 2003). This concept is foundational to our emerging understanding of the interaction between rapid adaptation and ecological dynamics: there is growing evidence that evolutionary changes are often as fast as ecological dynamics and can accelerate or counteract ecological changes. If the interacting species play a central role in shaping ecological communities or in material and energy transport through the ecosystem, then the effects of rapid evolution extend beyond populations to ecosystem dynamics. As such, as interacting species evolve at the molecular level there is the potential for changes at the highest levels of biological organization. A natural question is, What genes or loci are responsible for these eco-evolutionary dynamics?

While an increasing number of studies demonstrate the importance of rapid adaptation in a variety of systems, we know exceedingly little about the underlying evolutionary genetics that drive such systems. Determining the genetic variants that underlie rapid adaptation (that affects ecological dynamics) will contribute to a deeper understanding of the details of the repeatability of evolution; the evolutionary history of the genes driving heritable variation; and the nature of the genotype-phenotype map. If, for example, evolution is not highly repeatable across populations because there is a very narrow bridge from genotype to phenotype then the potential for evolutionary effects on ecological dynamics may be very narrow. In contrast, if the genotype-phenotype map is very broad—that is, if many different genetic solutions produce a necessary phenotype (sometimes termed *accessibility*)—then there may be high evolutionary repeatability and the role of rapid adaptation may be more generally important in shaping ecological dynamics. If we examine the history of the underlying molecular evolution, do we find a few, limited origins of ecologically important variants, or do mutations in the same genes (or even at the same nucleotide) arise and spread in multiple, independent lineages.

At this writing I am aware of only one study—an unpublished Master's thesis—that explicitly identifies a locus whose allelic variants extend to ecological communities. Jones (2003) used crosses of the cottonwood *Populus fremonttii* to map a locus affecting leaf tannin concentrations, a trait that drives the abundance and composition of canopy arthropod communities (cited in Whitham et al. 2006). This is a very interesting result, and is tempered only by the fact that the trait in question is near-Mendelian in its inheritance pattern. In contrast, many, if not most, ecological interactions involve complex quantitative traits; determining the genes and loci underlying variation in such

traits is a hard problem, and combining the difficulties of such research with the challenges of eco-evolutionary research makes for a "double-plus hard" problem.

An alternative to trying to solve the ecology, phenotypic evolution, and genetics of eco-evolutionary dynamics simultaneously is to build on established systems, taking one or more pieces of information as established "facts" and focusing on the complementary aspects. One option could be to start from well-developed model genetic species and explore the eco-evolutionary dynamics of such species. Unfortunately, very little is known about the natural ecology of classical model genetic species, and therefore an expansive research program would be necessary to establish if these species play significant ecological roles in the communities where they are found. Another option is to start from an ecological model species and employ contemporary –omics methods to quickly advance our understanding of the molecular basis of trait variation. While the ecological relevance of model genetic species is unknown and the search may prove fruitless, the importance of genetic variants to shaping the phenotypic diversity of model ecological species is assured—if there is heritable variation then there are underlying genetic variants to be discovered. Focusing on the development of genetic and genomic knowledge of model ecological species is the logical conclusion.

My thesis is that *Modern genomic, transcriptomic, and bioinformatic tools can be used to rapidly develop our understanding of the molecular basis of variation in traits that mediate eco-evolutionary dynamics*. I have focused on the common waterflea, *Daphnia pulex* (species complex), which has served as a model ecological species for over a century and which has been shown to exhibit rapid adaptation with concomitant effects on communities and ecosystems. In the first two chapters of this Dissertation, I identify suites of expressed genes, single-nucleotide polymorphisms (SNPs), promoters,

Gene Ontology terms, and gene networks associated with variation in 14 life history traits (Chapter 1) and 18 behavioral traits (Chapter 2) known to be important in shaping *Daphnia*'s ecological niche. The core results from this effort are extended in Chapters 3 and 4 to investigate the molecular basis of variation in the offspring size-number tradeoff, which is a fundamental tradeoff common across multicellular organisms; and to identify for *Daphnia*, which is a keystone ecological species, candidate "keystone gene networks" that are central to shaping its ecological requirements and impacts. Together these chapters support my thesis and the dissertation provides a significant and relevant contribution to the development of this emerging genomic model species that has long been the focus of ecological and evolutionary inquiry.

Chapter 1: Loci, Genes, and Gene Networks Associated with Life History Variation in a Model Ecological Organism, *Daphnia pulex* (complex)

Identifying the molecular basis of heritable variation provides insight into the underlying mechanisms generating phenotypic variation and the evolutionary history of organismal traits. Variation in life history traits is of central importance to ecological and evolutionary dynamics, and contemporary genomic tools permit studies of the basis of this variation in non-genetic model organisms. I used high density genotyping, RNA-seq gene expression assays, and detailed phenotyping of fourteen ecologically important life history traits in a wild-caught panel of 32 *Daphnia pulex* clones, to explore the molecular basis of trait variation in a model ecological species. I found extensive phenotypic and heritable genetic variation ($0.53 < H^2 < 0.76$) in the panel, and accordingly identify 75-261 genes—organized in 3-6 coexpression modules—associated with genetic variation in each trait. The trait-related coexpression modules possess well-supported promoter motifs, and in conjunction with marker variation at trans- loci, suggest a relatively small number of important expression regulators. I further identify a candidate genetic network with SNPs in eight known transcriptional regulators, and dozens of differentially expressed genes, associated with trait variation. The gene-trait associations include numerous un-annotated genes, but also support several a priori hypotheses, including an ecdysone-induced protein and several Gene Ontology pathways. The loci, genes, and coexpression modules associated with size and fecundity variation are candidates for the molecular mechanisms that shape *Daphnia* eco-evolutionary dynamics.

INTRODUCTION

Determining the relationship between genetic, trait, and ecological variation is a key goal of contemporary biological research. Investigating the relationship between genotype and phenotype—including genotype x environment interactions—is the domain of genetics, which rarely extends to higher levels of organization such as populations and communities. A complement at higher levels of biological organization is the field of trait-based ecology, which focuses on the relationship between phenotype and ecological processes such as population dynamics and community assembly (McGill et al. 2006; Ackerly & Cornwell 2007; Messier et al. 2010; Webb et al. 2010). A key step forward is to solidify these cross-hierarchy links in ecologically well-studied settings, but this has not yet been accomplished. The main reasons for this shortcoming are two-fold: we know very little about the ecology, and the ecological context of evolution, of most model genetic organisms; and genetic resources are sparse for model ecological organisms (Feder & Mitchell-Olds 2003; Tollrian & Leese 2010; Wheat 2010; Miner et al. 2012, 2012). Therefore, an ideal system with which to approach such a problem is one for which we possess both extensive genetic resources and ecological knowledge.

With the recent genome sequencing of the waterflea, *Daphnia pulex* (hereafter *Daphnia*), we can integrate information across levels of organization, from genetic sequence to ecologically important traits for an organism that is generally considered a keystone species (Carpenter & Kitchell 1996; Leibold 1996). The *Daphnia* genome is approximately 227MB in length and is characterized by numerous tandem duplications with rapid divergence of expression patterns (Colbourne et al. 2011). A century of ecological research has shown that *Daphnia* are central players in aquatic communities by acting as a key link between producers and carnivores; by controlling plant biomass

and production; and by their effects on nutrient cycling in lakes (see Lampert [2006] for a brief review). Arguably, we know the factors describing the requirement and impact niches (Chase & Leibold 2003) of *Daphnia* as well as, if not better than, any other species. Furthermore, we know that rapid adaptation in *Daphnia* can have dramatic effects on ecological dynamics (Hairston et al. 1999, 2005; Cousyn et al. 2001; Duffy et al. 2009; Ellner et al. 2011). The nexus of ecological and evolutionary knowledge with genomic tools for a single species enables linking the levels of biological organization in a way not possible for classical model genetic species and other model ecological species.

Several life history traits are central to shaping ecological and evolutionary dynamics, both in *Daphnia* and more generally. To provide context, I briefly review the general relevance of size, growth, and fecundity to ecological variation, including the specific connections to *Daphnia* biology. Next, I set out several a priori hypotheses concerning the genes and biological pathways I might expect, based on the literature, to be associated with variation in the life history traits examined. Finally, I define the goals of the present study—which are largely descriptive—before moving on to the results.

The ecological importance of size and reproduction traits

Body size, growth, and fecundity are among the ecologically most-important traits for a wide array of taxa. Few traits are thought to affect ecological dynamics more than body size (Peters 1986; Loreau 2010): metabolism scales with body size and is predictive of macroecological patterns (Enquist et al. 1999; West et al. 1999); body size limits the upper size of food an individual can consume (e.g., Bogdan & Gilbert 1984), and body size is intimately tied to predation susceptibility (Abrams & RoI 1996). Size is also strongly correlated with the number and size of offspring produced (Roff 1992), which is

a vital aspect of fitness and population growth. Individuals who tend to grow faster are larger as adults and tend to require more resources than individuals who are smaller. As a result, the nutritional requirements are greater, and the impacts on resource availability are greater, for larger individuals (Kreutzer & Lampert 1999).

These (and many other) trait:ecology mappings have been examined in *Daphnia*. Patterns of cladoceran body size, in *Daphnia* in particular, formed the empirical basis for Brooks and Dodson's influential size efficiency hypothesis (Brooks & Dodson 1965). They argued that larger species such as *Daphnia* are better competitors and should dominate any given community except in the face of vertebrate predators. While true in some cases, there are many caveats to the general pattern (Hart & Bychek 2010), including the fact that competitive outcomes with respect to body size are conditional on food quantity and quality: larger *Daphnia* have an advantage when algae is abundant and high-quality, but smaller organisms or individuals gain the upper hand when algae density or quality is low (Goulden et al. 1982; Tessier & Goulden 1982).

While size confers certain competitive advantages and disadvantages, it also shapes predation risk. A direct effect is that large *Daphnia* are more visible and therefore more susceptible to predation by vertebrate predators (Wright & O'Brien 1982; Wetterer 1989; Cernỳ & Bytel 1991). Rapid evolution—within a single season, and spanning just a few generations—of body size in response to seasonal changes in fish predation regimes has been shown in *Daphnia* (Tessier et al. 1992; Wolinska et al. 2007). In contrast to the interaction of size and vertebrate predators, small *Daphnia* are susceptible to predation by invertebrates, including the well-studied effects of the midge larvae (*Chaoborus* spp.) and copepods, which are unable to effectively handle larger prey (Spitze 1991; Gliwicz & Umana 1994; Beckerman et al. 2010; Dennis et al. 2011).

Furthermore, vertebrate and invertebrate predator regimes are not independent of one-another (Leibold & Tessier 1991; Beckerman et al. 2010), nor are they independent of other ecological processes such as competitive interactions (Gliwicz & Wrzosek 2008). Body size and its associated traits are thus key traits that mediate the relative susceptibility of *Daphnia* to different predators, and to be efficient resource exploiters in different habitats.

As competition and predation impose selection on *Daphnia* body size, correlated traits are also affected. For example, because reproductive output is often strongly correlated with maternal size, we expect that selection regimes favoring smaller *Daphnia* will also results in fewer and/or smaller offspring and affects influences ecological dynamics (Tessier et al. 1992). The size-efficiency hypothesis suggests that larger offspring have greater competitive ability when food is scarce, but more, smaller offspring provide a numerical advantage when food is abundant; Tessier and Goulden (1982) refined the hypothesis to state that larger individuals are at an advantage when food availability fluctuates extensively. Larger offspring tend to have higher growth rates, however, which requires higher nutrient concentrations to support growth (in particular, phosphate; Sterner & Elser 2002; Elser et al. 2003). The nutrient environment in which *Daphnia* grow and reproduce is shaped by abiotic and biotic factors, e.g., shading alters algal stoichiometry to increase the amount of phosphorus relative to carbon, which interacts with genotype to alter growth rate and body size (Main et al. 1997; Urabe et al. 2002; Hall et al. 2004; Weider et al. 2005; Shimizu & Urabe 2007).

Expectations

While the genomic resources for *Daphnia* are a recent development, we can make several predictions about the genes and pathways expected to be associated with life history trait variation. Genome-wide expression and genotyping studies facilitate discovery of novel gene-trait associations (Bruggeman & Westerhoff 2007), but post-hoc explanations of patterns are weaker than tests of a priori hypotheses. Decades of classic molecular, reverse-, and forward-genetic studies provide guidance as to the pathways and genes expected to be related to trait variation; before I analyzed any SNP-trait or expression-trait data for *Daphnia*, I searched the existing literature for candidate genes and pathways that might allow us to make a priori hypotheses. Recovering expression or genotype variation in these pathways, related to trait variation, provides additional support for the statistical inference.

Insulin and insulin-like signaling pathways are commonly related to size- and growth-related traits in numerous model species (Stocker & Hafen 2000; Brogiolo et al. 2001; Fujiwara et al. 2002; Boucher et al. 2010). Ecdysone-activated proteins, which regulate arthropod molting and were the top candidates in experimental evolution of fly body size (Turner et al. 2011), are found in *Daphnia* (e.g., nucleolar protein c7b, which is a homolog of the *Drosophila* gene *mustard*). FOXO, a forkhead transcription factor, is central to insulin signaling and stress response, the latter of which was found to be a top body size-related pathway using flies from the *Drosophila* Genetic Reference Panel (DGRP; Jumbo-Lucioni et al. 2010; Mackay et al. 2012). Neuronal control of size—through related behaviors such as movement patterns and feeding rate—has been demonstrated with *C. elegans* (egl-4, a cGMP-dependent kinase (Fujiwara et al. 2002))

and *Drosophila* (short neuropeptide F, a protein precursor thought to be central to chemosensation; Lee et al. 2004). Although some of these examples are very specific, they provide guidance on what to expect when exploring the genomics of size and growth variation in a newly developing model species such as *Daphnia*.

Less information is available about the genes and pathways underlying variation in the number of offspring or time between broods (clutches). The Gene Ontology categories for genes best-correlated to absolute fitness in flies from the *Drosophila* Genetic Reference Panel include proteolysis, signal transduction, and defense/immune response (Ayroles et al. 2009). There are numerous vitellogenin-like genes in the *Daphnia* genome, and contaminant stressors have been shown to affect the expression of vitellogenin genes and the levels of the vitellogenin antagonist, juvenile hormone in *Daphnia* magna (Heckmann et al. 2008). The copy number of yolk protein genes (*yp-1*, *yp-2*, and *yp-3*) is positively correlated with egg production in *Drosophila* (Bownes et al. 1991), but there are no clear *Daphnia* homologs (the closest BLAST hits are to a pancreatic triacylglycerol lipase). Of note, the yp genes share sequence similarity with vertebrate lipases, which is thought to underlie the ability to store various steroids used in developmental signaling (Bownes 1992). I expect to find support for some or all of these genes, plus new candidates, by association with variation in *Daphnia* fecundity traits.

Long gene lists are often of less utility than general terms associated with the function of groups of genes for understanding the biology of variation in quantitative traits. After considering a combination of the literature and basic biology, I developed a matrix of biological terms, guided by the Gene Ontology (Ashburner et al. 2000) and KEGG (Kanehisa & Goto 2000) frameworks, expected to be enriched for groups of the growth and fecundity traits (Table 1.1). These relationships are straight-forward, e.g., we

anticipate enrichment for metabolism and genetic information processing groups of terms because of the need to convert materials into biomass, and several signaling and cell cycle-related terms because of their role in development and growth.

Present goal

Given the general ecological importance of size, growth, and fecundity; the well-established role variation in these traits plays in shaping *Daphnia* ecological and evolutionary dynamics; and a desire to understand the connections across the genetic, trait, and ecological levels of organization, I investigated the molecular basis of life history variation in a panel of wild-collected *Daphnia*. I quantified variation in fourteen growth and fecundity traits, used RNA-seq to quantify constitutive gene expression variation across the panel, and genotyped each clone at an average of ~3 million loci. I then integrated this data to provide estimates of the relationships between genetic, expression, and phenotypic variation. The results indicate a relatively small number of coexpressed gene modules are associated with variation in these traits, provide novel hypotheses to be explored as the field of ecological genomics expands, and establish several candidates for genes underlying the interface of eco-evolutionary dynamics.

RESULTS

Phenotypic variation

There was substantial variation across the panel of 32 clones for all 14 traits that I measured. First, the difference between the smallest and largest clones was approximately four-fold for juvenile mass and three-fold for adult mass (Figures 1.1A-C). The differences were smaller for *Daphnia* length and depth (Figures 1.1D-E), but of

similar magnitude for growth rate (Figure 1.1F). The variation in body sizes translated to substantial variation in the number of offspring, with up to a 9-fold difference in brood size (Figures 1.2A-C). In addition, there was up to a three-fold difference in interbrood period (Figures 1.2D-F). Clone-wise summary statistics are provided in SI Table 1.1.

Broad-sense heritability varied from 0.53-0.76 for the growth and fecundity traits (Table 1.2), but the traits are not independent. Phenotypic correlations (Figure 1.3, above-diagonal) were generally weaker than genetic correlations (Figure 1.3, below diagonal), and each tended to be in the directions expected: adult mass was positively correlated with external measurements, the number of offspring per brood, time to first reproduction and the interbrood period. Adult mass was negatively correlated with the mass of offspring in the first brood: larger females invested in more offspring rather than allocating resources to larger offspring. Although the relationship was weaker, the number of offspring per brood was negatively correlated with the size of those offspring. One prominent exception to the expected correlations was the negative relationship between offspring size and growth rate: smaller neonates grew faster than larger neonates. The trait with the lowest heritability—number of offspring in the third brood—was also the trait with the largest difference between phenotypic and genetic correlation: environmental variation reduced the correlation with adult mass to 0.22. Similarly, environmental variation increased the negative correlation between the size of first-brood offspring and number of offspring in the second brood. The inter-trait correlations are further emphasized by a principal components decomposition of the phenotype data (i.e., all replicates), where the first 5 axes explain 97% of variation among all fourteen traits.

Genotypic variation

Sequencing from 2b-RAD and RNA-seq libraries resulted in an average 2.2 x 10^6 raw reads per clone, which, when aligned to the *Daphnia* reference, resulted in 5.5 x 10^6 genotyped base positions across the population and an average of 4.66 x 10^6 loci genotyped per clone (s.d. = 1.2 x 10^6). There were 6338 loci polymorphic in the panel and typed in >75% of clones; heterozygosity rate was low, with a mean across clones of 0.14%. Mean indel and substitution rates estimated during read mapping were 2.7 x 10^{-4} and 0.022, respectively. I observed few patterns of genome-wide genetic variation; for example, mean nucleotide diversity in 50kb sliding windows was relatively constant (Figure 1.4). Population structure is relatively low in the panel, and linkage disequilibrium appeared to decline relatively quickly, to background levels by about 200bp (see Chapter 2).

Gene expression variation

As expected given the wide variation in size, growth, and fecundity traits, I uncovered substantial variation in constitutive gene expression across the 32 clones. Although > 30,000 *Daphnia* genome features (i.e., genes) possessed at least one mapped read, subsequent analyses are based on a subset of 15,600 genes with mean expression of 5 reads per million mapped. Approximately 45% (n = 6434) of genes were differentially expressed (DEGs) at $p < 0.05$ after Benjamini-Hochberg FDR control at a 1% rate, using a generalized linear model with quasipoisson errors (see Methods). As expected, DEGs possess a significantly mean higher broad-sense heritability ($H^2 = 0.657$) compared to all expressed genes ($H^2 = 0.508$, $p < 2.2e\text{-}16$; SI Figure 1). The degree of expression

differences varied from 1.6- to 7000-fold, due only to genetic differences between clones in a common garden environment.

Global gene expression was highly modular, with 24-27 distinct coexpression modules recovered across a wide range of parameter conditions (SI Figure 1.2). The genes within each module were associated with 1-17 novel promoter motifs (1000bp upstream and 200bp downstream of the transcription start site [SI Table 1.2]) based on word enrichment as determined with XXmotif (Hartmann et al. 2013). Each module was enriched for 6-97 Gene Ontology (GO) terms, ranging from generic "biological process" to more detailed "hydrolase activity" (SI Table 1.3).

Linking genotypic, expression, and trait variation

I identified 14 SNPs (nominal $p < 1e\text{-}5$) and an average of 156 DEGs ($p < 0.05$ and FDR control at 1%) whose variation was tightly correlated with variation in each of the fourteen growth and fecundity traits (Table 1.3, SI Table 1.4). Two gustatory receptors (hxAUG25s1441g78t1, hxJGI_V11_235732) were differentially expressed and associated with second and third brood sizes, but no insulin receptor or peptide-encoding genes were associated with growth and fecundity traits. One SNP (scaffold 17:1051920) was located in an ecdysone-induced protein gene, and one differentially expressed ecdysone receptor (hxAUG26res30g96t1, 20-hydroxy-ecdysone receptor 20e) was associated with variation in offspring mass and adult tailspine length. Several lipases are differentially expressed and strongly correlated with number of offspring, time between broods, and adult tailspine length. There are currently 26 vitellogenin-associated genes annotated in the *Daphnia* genome; expression of only one vitellogenin precursor,

hxJGI_V11_307854, was associated with variation in adult body length, but not brood size, offspring size, or time between broods.

I recovered 3-6 modules of coexpressed DEGs per trait (Table 1.3), drawn from an average of 4.3 global expression modules (range 2-8; SI Figure 1.3). There were substantial differences in the architecture of trait-specific coexpression networks. For example, while there is a similar number of genes associated with adult mass (Figure 1.5A) and third brood offspring mass (Figure 1.5B), the adult mass network has more "hub" genes (i.e., with high betweenness-centrality, a metric of the node's importance in connecting the network) than the offspring mass network. The adult mass gene with the highest betweenness centrality is *Partner of bursicon*, a precursor for a neurohormone involved in molting (Mendive et al. 2005); the two hub genes of the offspring mass network encode ERK-a and Rab-32 proteins, members of the MAPK/ERK signaling pathway (Seger & Krebs 1995).

Although 60% of genes associated with *Daphnia* growth and fecundity variation possess no functional annotation, I identified between seven and 79 GO terms enriched for each growth and fecundity related trait (SI Table 1.5). Across all fourteen traits, GO terms associated with gene expression regulation, protein transport, metabolism, cell proliferation, and signaling were among the top enriched Biological Process (BP) groups (Figure 1.6A). Correspondingly, nucleic acid binding, transporter activity, macromolecule binding, and receptor activity were among the groups of Molecular Function terms enriched (SI Figure 1.4). Traits characteristic of adult *Daphnia* (i.e., size and growth rate characteristics) were enriched for BP terms related to protein transport, cell cycle and proliferation, and growth (Figure 1.6B), while offspring-related traits were

enriched for gene expression regulation, developmental processes, signaling, and cell death-associated BP terms (Figure 1.6C).

One mechanism driving gene coexpression is shared expression regulators, and by extension, shared recognition sequences in promoter regions. I identified between one and 17 promoter motifs enriched in the promoters of genes associated with each trait (SI Table 1.6). The number of motifs recovered per trait was proportional to the number of genes in each coexpression module considered, and support for each motif ranged from E-value < $1e^{-3}$ to $1.4e^{-26}$. Motifs containing TGC repeats, such as Motif 1 (MTGCTGCTGCTGCTGYY) of first-brood offspring mass "turquoise" module, were associated with half of all growth and fecundity traits, suggesting that it may be a target of a general transcriptional regulator associated with size and reproduction; there was no strong sequence similarity with known *Drosophila* motifs, however. Several motifs with long cytosine repeats, such as the motif with the lowest E-value (CYCCCCCCCCCCYHYHB, $1.43e^{-26}$), are highly similar to the *Drosophila* motif target of CG7368, an unnamed transcription factor associated with phagocytosis (FlyBase 2007). The *Daphnia* gene with the highest sequence similarity to CG7368, ZFP-ZMS1 (hxAUG26us24g213t1), was differentially expressed but not strongly associated with variation in any of the traits.

An integrated network hypothesis

I intersected the DEGs strongly associated with SNP variation and trait variation to identify a systems genetic—i.e., spanning from genetic to expression to organismal variation—hypothesis for several growth and fecundity traits (Figure 1.7). This hypothesis is restricted to SNPs occurring in genes with transcription regulation

annotations; although any of the SNPs associated with gene expression or trait variation (SI Table 1.7) may be causal or linked to causal variants, the annotations associated with these SNPs suggests a plausible regulatory mechanism. One of the markers is within an ecdysone-induced protein (hxAUG26us17g279t1) that is not differentially expressed, but the allelic variants strongly predict ($p < 1e^{-7}$) the expression levels of 27 DEGs. The MAPK ERK-a and PA2G4 markers are both associated with cell proliferation through different pathways (Seger and Krebs 1995), and are linked to time to first brood through associations with four DEGs.

DISCUSSION

One goal of contemporary biology is to predict the drivers of variation between levels of organization, from genetic sequence to ecosystems. The importance of the core life history traits of size, growth, and fecundity in shaping ecological dynamics has long been hypothesized, is encompassed by ecological theory, and we have empirical support of their importance in a few systems, including the ecological model organism, *Daphnia pulex*. The recent sequencing of the *Daphnia* genome means that we can begin to relate genetic, trait, and ecological variation, as well as identify the genes underlying rapid evolution and its ecological consequences. In this paper I provide a first pass at the map from genetic to trait variation in traits known to affect *Daphnia* ecology, given a panel of wild-caught *Daphnia* clones. I find substantial variation at three levels of biological organization—genotype, gene expression, and organismal phenotype—while linking the levels and generating novel hypotheses to test as the *Daphnia* system matures for genomic research.

The amount of observed natural phenotypic variation was not surprising given the variety of habitats—abiotic, competitive, and predation regimes—from which the panel was collected. Previous studies of *Daphnia* that measured organismal traits also recovered substantial levels of variation (e.g., Lynch 1984; Leibold & Tessier 1991; Spitze et al. 1991; Pfrender & Lynch 2000) among clones, but often were more focused on interspecific differences. Some trait correlations qualitatively matched those of previous studies; for example, Spitze and colleagues (1991) found a strong positive correlation between growth rate and reproductive output, as I found. However, while I found effectively no correlation between time to first brood and brood size, Spitze and colleagues recovered a negative correlation (-0.08 to -0.69, depending on the brood). I did not observe sign change between genetic and phenotypic correlations for any trait pair, and the degree of change was within expectations given heritabilities (Lynch & Walsh 1998).

Our genome-wide genotyping offers one of the first population genomic examinations of the D. *pulex* group. While the panel is relatively small, I sampled individuals of all three ecotypes (two lake, fourteen shaded-pond, 16 sunny-pond) and found extensive genome-wide genetic variation with relatively little population structure. Coupled with the continuously distributed variation among the fourteen traits, this result supports reinforces other recent findings (Adamowicz et al. 2009; Heier & Dudycha 2009; Cristescu et al. 2012) suggesting continued gene flow between the ecologically isolated *D. pulex* and *D. pulicaria*, through hybrid back-crosses. Future population genomic studies with larger samples, and from across a larger geographic area, will shed light on the enigma that is the D. *pulex* group.

Our analysis recovered a highly structured and modular transcriptome in *Daphnia*. This finding reflects my expectations given the mechanisms of expression control (i.e., relatively few transcriptional regulators with many targets) and the results of similar studies (Stuart et al. 2003; Ayroles et al. 2009). Enrichment for a relatively small number of Gene Ontology terms per module, and the presence of well-supported promoter motifs among coexpressed genes, further supports the cohesion of the modules. Putative functional annotations may be transferred to currently un-annotated DEGs (~ 66% of DEGs) given the patterns of coexpression and shared promoters with functionally annotated genes (Škunca et al. 2012), advancing our knowledge of *Daphnia* molecular biology while generating hypotheses that can be tested by more focused approaches. Such extension is, however, beyond the scope of the present work.

In addition to support for several a priori expectations—from ecdysone-activated genes and gustatory receptor gene expression, to enrichment for a variety of biological terms—I also identified hundreds of candidate genes associated with variation in the fourteen *Daphnia* growth and fecundity traits. This magnitude of discovery is expected from genome-wide approaches and provides a foundation for novel hypothesis-driven molecular research with *Daphnia*. For example, the fact that >60% of DEGs associated with *Daphnia* growth and fecundity variation have no functional annotation provides fertile ground for new discoveries and insights into the functional diversification of genes. Other results are better-known: the central role of *Partner of bursicon*, ERK-a, and Rab-32 to *Daphnia* size variation networks captures these genes' known roles in development and cell proliferation in *Drosophila*.

The coexpression network analyses add to our collective knowledge about the inherent modularity of biological systems. Modularity—sets of interactions that are

stronger within than between groups—is a common feature of biological systems, from neural networks to genetic architecture (Alon 2003; Barabasi & Oltvai 2004). Modularity itself may be a target of selection by decreasing interference between the modules (Espinosa-Soto & Wagner 2010), and intermediate levels of pleiotropy across modules should facilitate evolvability (Hansen 2003, 2006). Among the coexpression networks associated with variation in *Daphnia* traits examined here I found substantial variation in the degree and details of modularity. For example, although offspring and adult mass are the same character from different life stages, differences in the number of modules (three vs. five) and number of high betweenness-centrality genes (two vs. five) associated with each trait suggests different evolutionary potentials. How these coexpression modules are remodeled (Ideker & Krogan 2012) across the developmental timeline, and in a variety of habitats, is an open question; the results provide a reference against which future research can be compared.

Variation in the GO terms enriched for the sets of genes associated with individual and groups of *Daphnia* growth and fecundity traits highlights similarities and differences in the underlying biological processes. Cell cycle, proliferation, and death are common between the groups of traits, but enrichment for protein transport and localization is restricted to adult size traits, while developmental processes are (as expected) enriched in offspring. I are not aware of similar analyses and comparisons in other species, but this observation suggests that quantifying numerous similar traits—rather than just one or two representatives of a trait class—has the potential to inform our understanding of the molecular basis of many small phenotypic distinctions (Houle 2009; Houle et al. 2010). Note, however, that inferences from biological term analysis need to

be tempered with the fact that most *Daphnia* genes are lineage-specific (Colbourne et al. 2011) and currently possess no functional annotation.

The goal of systems genetic approaches is to identify and understand the functional relationships between genetic variants; the genes whose expression (and post-transcriptional modifications) are affected by the genetic variants and environmental causes; and ultimately phenotypic variation (Loewe 2009; Nadeau & Dudley 2011). The combination of extensive genotype, expression, and phenotype data allowed us to generate a systems genetic hypothesis linking variation across these three levels of organization for *Daphnia* growth and fecundity. The relatively small size of the panel and lack of controlled crosses precluded a full probabilistic analysis of the network, and the focus in particular on genetic variants in "known" transcriptional regulators is cautious: for example, by requiring transitivity from genotype to organismal trait (i.e., markers associated with DEGs and one or more traits, with the DEGs also correlated with variation in the same trait) we restrict the core hypothesis (Figure 1.7) to eight markers and three traits. Many other loci and genes likely underlie growth and fecundity variation in *Daphnia*, but the elements of this network are supported in various analyses and provide hypotheses to be tested. For example, the systems hypothesis suggests a major role for an ecdysone-induced protein on scaffold 17; as a major regulator of arthropod molting (Ishimoto & Kitamoto 2010), the role of ecdysone (or rather, the targets of ecdysone) is a strong candidate for general regulation of *Daphnia* growth and fecundity. Future work that perturbs the *Daphnia* system—through crosses, RNAi (Kato et al. 2011), plasmid integration (Kato et al. 2012), environmental manipulations (Jeyasingh et al. 2011), and other methods—will help expand upon and test the present hypothesis.

While the present work is a distinct advance for *Daphnia* genomics, the limitations of the approach used here must be recognized. First, the expression and phenotypic data were collected in a single, benign common garden setting, and *Daphnia* is well-known for its extensive phenotypic plasticity (e.g., Spitze 1992; Tollrian 1993; Bernot et al. 2006; Jean-Christophe et al. 2011). These results provide a baseline against which future work, under a variety of ecologically interesting and realistic conditions, can be compared. Second, I used whole *Daphnia* for RNA collection, but tissue-specific expression differences are a well-known phenomenon among many organisms and genes (Wray et al. 2003). Future research that aims to isolate expression to particular *Daphnia* tissues will certainly refine our understanding of the relationship between genotype, expression, and morphological variation.

There are two main implications of the present research for our collective understanding of *Daphnia* ecology and evolution. First, this is one of the few examples, if not the only example, of ecological genomics (Antonovics 2003; Stearns & Magwene 2003) applied to an organism whose community ecological context is very well-studied. Because the implications for size and fecundity variation in *Daphnia* are known to extend both up and down trophic levels, identifying loci and genes affecting these traits drives at causes of the extended phenotype (Dawkins 1982). That is, if the genotype at scaffold_17:1051920 (a T/C or T polymorphism) affects *Daphnia* size (through numerous intermediary genes), then it is therefore predictive of the effects on phytoplankton communities through grazing pressure, and susceptibility to predation by vertebrate and invertebrate predators. This example and the hundreds of others discovered in this work provide numerous hypotheses to be tested as explanations of organismal and community variation.

Second, it has become apparent over the past decade that rapid evolution—changes occurring on the scale of just a few generations—can have dramatic ecological implications (Thompson 1998; Cousyn et al. 2001; Yoshida et al. 2003; Hairston et al. 2005). *Daphnia* has been a model system for examining eco-evolutionary dynamics of disease (Duffy & Sivars-Becker 2007), predation (Fisk et al. 2007; Scoville & Pfrender 2010), and eutrophication (Hairston et al. 1999), and in each of these cases the basic life history traits studied here play a central role in shaping the ecological interactions. An outstanding issue is identifying the loci underlying evolutionary change of ecological importance. For example, phenotype data alone cannot answer whether the same loci are involved in parallel bouts of adaptation, or if there are multiple, unique avenues of adaptation (see, e.g., Arendt & Reznick 2008; Rosenblum et al. 2010). Because the panel was collected from natural populations, the results provide candidates for the loci that may be responsible for rapid, ecologically important evolution in *Daphnia*.

Here I have identified genetic and gene expression variants associated with variation in life history traits of the model ecological organism, *Daphnia pulex*. In addition to recovering several expected gene/pathway relationships, the analyses uncovered numerous novel gene-trait relationships that form the basis for future hypothesis-driven research. These results are an important first step for understanding the molecular basis of variation in *Daphnia* growth and fecundity—traits for which the impacts of variation extend across communities—and are candidates for the molecular basis of ecologically important adaptation.

METHODS

Clone collections and maintenance

Nominal *Daphnia pulex*, *D. pulicaria*, and hybrid clones were collected from a variety of waterbodies in the area surrounding Kellogg Biological Station, Michigan, USA, in June 2009. The panel includes clones from the mesocosm experiments of Pantel et al. (2011). Individuals were isolated and cultures started in Austin, TX, from a single female; after initial mortality of isolates and clone losses during the assay period, I obtained 32 unique lineages. All working cultures were maintained in ADaM media (Klüttgen et al. 1994) in an environmental chamber at 20°C with 16:8 L:D cycles. They were fed daily 1-2ml Shellfish Diet (Reed Mariculture, Campbell, California), at $2x10^6$ cells per milliliter, regularly supplemented with live *Scenedesmus acutus*.

Phenotypic assays

Daphnia are model organisms for studying maternal effects (Alekseev & Lampert 2001; Mitchell & Read 2005), but such effects are not the focus of this work. To minimize maternal and grand-maternal effects on size, growth, and fecundity estimates, each replicate was reared through two generations of single-individual breeding before assaying the focal generation. All individuals in the assays were raised in 100ml cups in ADaM media and fed 1ml shellfish diet each day, supplemented with 4x105 cells of *S. acutus* every-other-day. Half of the media was replaced in each cup every 3d, and each cup was thoroughly cleaned and all media replaced each week. Each cup was checked daily for the presence of offspring and for mortality. If offspring were present, then they were counted and 2-4 placed into a new cup with fresh media and food. These individuals were randomly culled to a single individual within 2d of their birth, but young

individual mortality required that >1 be retained initially. If the focal individual had died at the check, then the replicate was started over at the grand-maternal generation with an individual taken from the working culture.

Once the focal generation had been reached, I recorded the dates and sizes of the first, second, and third broods. All offspring of the first and third broods were collected and stored in a 1.5ml tissue tube in 95% ethanol in a -10oC freezer. The mother was collected and stored in the same manner at the release of her third brood. At regular intervals, I removed samples from storage for further measurement. Adult body length to the base of the tail spine, body depth at the deepest point, and tail spine length were measured to the nearest 0.1mm under a 10-40x dissecting microscope. Adults and 1-15 juveniles were then placed in pre-tared aluminum mini weigh boats and dried for 48-72h at 60°C prior to weighing. *Daphnia* mass was measured to the nearest 0.1μg on a Sartorius ultramicrobalance placed in a closed room on a stabilizing marble bench. (Note that I rounded all measurements to the nearest 1μg for analysis.) Growth rate was calculated as the difference between log(mean neonate mass) and log(adult mass) divided by the number of days between the birth and the third brood. Adults were collected after weighing, and stored at -10°C for stoichiometric analysis. Percent phosphorus of each adult was established by measuring absorbance spectrophotometrically at 850nm.

Given the phenotypic data for three replicates of each clone, I calculated genotypic trait values by ANOVA of form trait ~ clone, and estimated heritability from the genetic variance ([mean square between - mean square error] / n) and total phenotypic variance (Lynch & Walsh 1998). Genetic correlations between traits was calculated from genotypic trait values and phenotypic correlations from all data.

Genomic samples

I measured constitutive gene expression for each clone as a starting point for interrogating the genotype-phenotype map in *Daphnia*. To do so, I raised three replicate sets of cultures of each clone for RNA sampling in the same environmental chamber, and using the same feeding regimen, as the working cultures. Each culture was maintained at a low density of 8-12 individuals per 150ml ADaM, for three generations. Even at this low density and high food provision, some clones appeared to be producing ephippia (resting stage eggs) whereas other clones were not. Collections were marked if there was any sign of ephippia production, which would likely alter gene expression profiles. Three individuals were collected from each replicate and each was stored in individual collection tubes to facilitate single-individual analysis of expression. That is, I collected a total nine individuals per clone. The collections took place on two adjacent days from 10:00-13:00h local time to minimize any circadian effects. Samples were placed immediately into a liquid nitrogen-filled Dewar, then transferred to and stored in a -80oC freezer.

After grinding by mortar and pestle in a liquid nitrogen bath, RNA was extracted from single individuals using Qiagen RNeasy kits (Qiagen, CA) per the manufacturer's instructions. RNA preparation for SOLiD sequencing followed the basic method of Meyer et al. (2011). In brief, this is a 3' tag RNA-seq method whereby fragmented RNA is reverse-transcribed to a cDNA library, amplified, and tagged with a SOLiD-ready barcode. Initial fragmentation was accomplished by a single 3-minute period at 95oC in a thermocycler; fragmentation was confirmed by gel analysis. After fragmentation, a cDNA library was created by reverse transcription using SuperScript II reverse transcriptase and a switching template primer. I amplified the cDNA library using

Titanium Taq and a thermocycle regimen of 5-min at 95C then 19 cycles at 95C (40s), 63C (1min), and 72C (1min). The PCR products were purified using a NucleoMag 96 cleanup kit, per manufacturer's instructions. I quantified DNA concentrations after cleanup using a NanoDrop spectrophotometer. Barcodes were ligated to each sample using the SOLiD multiplex P1 oligo, 1uM barcode oligo, Titanium Taq, and a amplification profile of four cycles at 95°C (40s), 63°C (1min), and 72°C (1min). After amplification, samples were run on a 1x TBE gel, and size-selected between 180-250bp using a low molecular weight ladder (NEB #N3233S). The cDNA in each gel slice was extracted by immersing the gel slice in nuclease-free water overnight at 4°C. The 96 prepped libraries were given to the University of Texas Genome Sequencing and Analysis Facility (UT GSAF) for sequencing on SOLiD 5500XL and v4 platforms, with a target of 2 million 50-bp reads per sample.

Single *Daphnia* individuals did not yield sufficient gDNA for RAD genotyping, so I pooled three individuals of each clone for extraction. Genomic DNA (gDNA) was extracted using Qiagen DNeasy extraction kits (Qiagen, CA) per manufacturer's instructions, with two exceptions: I did not vortex samples, in order to ensure that gDNA remained intact before *AlfI* digestion, and I completed the final rinse with 40μl nuclease-free water rather than 100μl to increase yield concentration. In brief, the 2b-RAD genotyping method (Wang et al. 2012) uses the *AlfI* restriction enzyme to digest the gDNA: SOLiD-system adaptors were ligated to the digested DNA and unique barcodes are then incorporated with the ligated products for each sample. The target constructs are 136bp in length and were extracted from electrophoretic gels. Samples of the 32 clones were prepared and sent to the UT GSAF for sequencing at a target of 1 million 50bp reads per sample.

Genotyping

I used the GATK (McKenna et al. 2010) two-phase genotyping process to call genotypes from the filtered SAM files. First, I used the IndelRealigner to correct for insertion and deletion errors inferred during mapping, then performed an initial genotyping pass with UnifiedGenotyper (using –ploidy 2, -glm BOTH, and –out_mode EMIT_ALL_CONFIDENT_SITES). From the first-pass genotyping I extracted variant sites in the 90th percentile of both genotype qualities (GQ) and QUAL scores as high-quality genotypes to be used in base quality score recalibration. After applying BaseRecalibrator I performed a second round of genotyping and extracted variants with GQ > 30 ($p < 0.001$). Subsequent analyses based on genotypic variation used a subset of loci in which ≥ 78% of clones (≥ 25) were genotyped.

Two risks of using RNA-seq derived reads for genotyping include allele-specific expression (Knight 2004) and RNA editing (Li et al. 2009b). To test if these potential sources of error were introducing significant bias, I genotyped all clones with DNA- and RNA-only data sources and compared genotype calls at overlapping sites: if either error source is prevalent, I expect to detect many more heterozygous loci from DNA-derived sequence. There were an average of 7122 overlapping loci per clone, and 10.3 differences on average, but both DNA- and RNA-derived genotypes possessed heterozygous calls when the alternate call was homozygous. Given the very low error rate (0.14%) and the fact that *Daphnia* genomics is a very young field, I opted to retain the joint genotype data with a very slightly higher false discovery risk.

Expression analysis

I used the classic SHRiMP output and the probcalc function for expression analysis, classifying mappings with normodds > 0.66 (odds-ratio ≥ 2 for the next-best mapping) as uniquely mapped. Ambiguously mapped reads were allocated to multiple genes in proportion to the support for the mapping and the proportion of uniquely mapped reads assigned to each gene. While allocation is required to reduce bias against multi-copy versus single-copy genes, it is known that allocating multimapped reads biases low-expressed genes high (Taub et al. 2010), however, because low-coverage genes (< 5 reads per million mapped) were excluded from analysis this source of bias is negligible in the expression data. Last, to achieve comparability between clones with different sequencing coverage, all expression levels were converted to reads per million mapped and rounded to the nearest integer.

I used a generalized linear model (GLM) with quasipoisson errors, log link function, and variance inflation estimated for each gene from the pooled data to test for differentially expressed genes (DEGs), with clone ID as the predictor. P-values were derived by likelihood ratio test against the null model, significance was set at $p < 0.05$, and Benjamini-Hochberg false discovery rate (FDR) control was exerted at a 1% rate.

Marker- and gene-trait associations

I tested for marker-trait variation using a linear model and set significance at a nominal p-value of $1e^{-5}$. Only markers with at least three clones possessing the minor allele (i.e., 9-12% minor allele frequency, depending on the number of clones typed at a locus) were used in this analysis.

I used distance correlation, which is not dependent on linear or monotonic relationships between variables (Székely et al. 2007; Székely & Rizzo 2009), to quantify the strength of association between organismal trait means and gene expression means. The p-values of distance correlations were derived from extensive bootstrapping of expression data, and FDR control exerted at the 1% level. Genes whose expression was significantly associated with variation in each trait were retained as initial candidates underlying variation.

Coexpression networks and candidate gene refinement

I used WGCNA (Langfelder & Horvath 2008) to (a) identify gene coexpression modules associated with each trait and (b) refine the list of candidate genes. I applied WGCNA to the residuals of trait:gene regressions to reduce the occurrence of false-positives; three regression models (linear, log-limited, and negative exponential) were tested for each trait-gene combination because non-linear relationships from distance correlation were detected, and the residuals from the model with the highest R^2 were used. Genes that did not cluster into a module (i.e., assigned the "gray" module of WGCNA) were removed from the candidate list as likely false-positives (Ayroles et al. 2009). Modules were identified by estimating the exponent required for a scale-free distribution, and module membership refined by adjusting reassignThreshold and mergeCutHeight to best-match the modules apparent in the heat map. Additional flags included pamStage=FALSE, TOMType="unsigned", TOMDenom="mean", and minModuleSize=8. Network figures were created in Cytoscape by exporting the WGCNA data for the best-connected.

After identifying the final candidate gene list for each trait I used the GOstats package for R (Falcon & Gentleman 2011) to test for Gene Ontology (GO) term enrichment. I defined gene universes for (a) all *Daphnia* genes with GO annotations and (b) *Daphnia* DEGs with functional annotations. Because relatively few *Daphnia* genes possess functional annotation (~25%), I set pvalueCutoff at 0.1. Enrichment was quantified both at the level of all candidate genes for a trait and module-wise for the WGCNA-defined modules associated with variation in a trait. I used REVIGO (Supek et al. 2011) for GO term enrichment visualization.

I used XXmotif (Hartmann et al. 2013) to identify candidate promoter motifs shared among the genes and modules associated with variation in each trait. Promoters were extracted from the *Daphnia* genome file from 1000bp upstream and 200bp downstream of each gene's transcription start site, both strands were searched for motifs, a background model of order 2 was employed, and a medium threshold was used for merging similar motifs. I retained motifs with $E < 0.1$, which corresponds to $E < 0.01$ because XXmotif estimates are biased high by approximately an order of magnitude for these sample sizes (see Hartmann Supp. Info. 1). I searched TOMTOM (Gupta et al. 2007) for known motifs in *Drosophila* similar to those recovered with XXmotif.

To identify candidate regulators of trait-specific coexpression modules, I mined (case-insensitive `grep`) module members with annotations for "transcriptional regulation", "kinase" and "MAPK", because the known role of TFs, kinases, and particularly mitogen-activated phosphate kinases in transcription regulation. The integrated network of Figure 1.7 was created by intersecting the (transcription-associated) marker-DEG, DEG-DEG, and DEG-trait data to establish relationships across levels of biological organization.

FIGURES

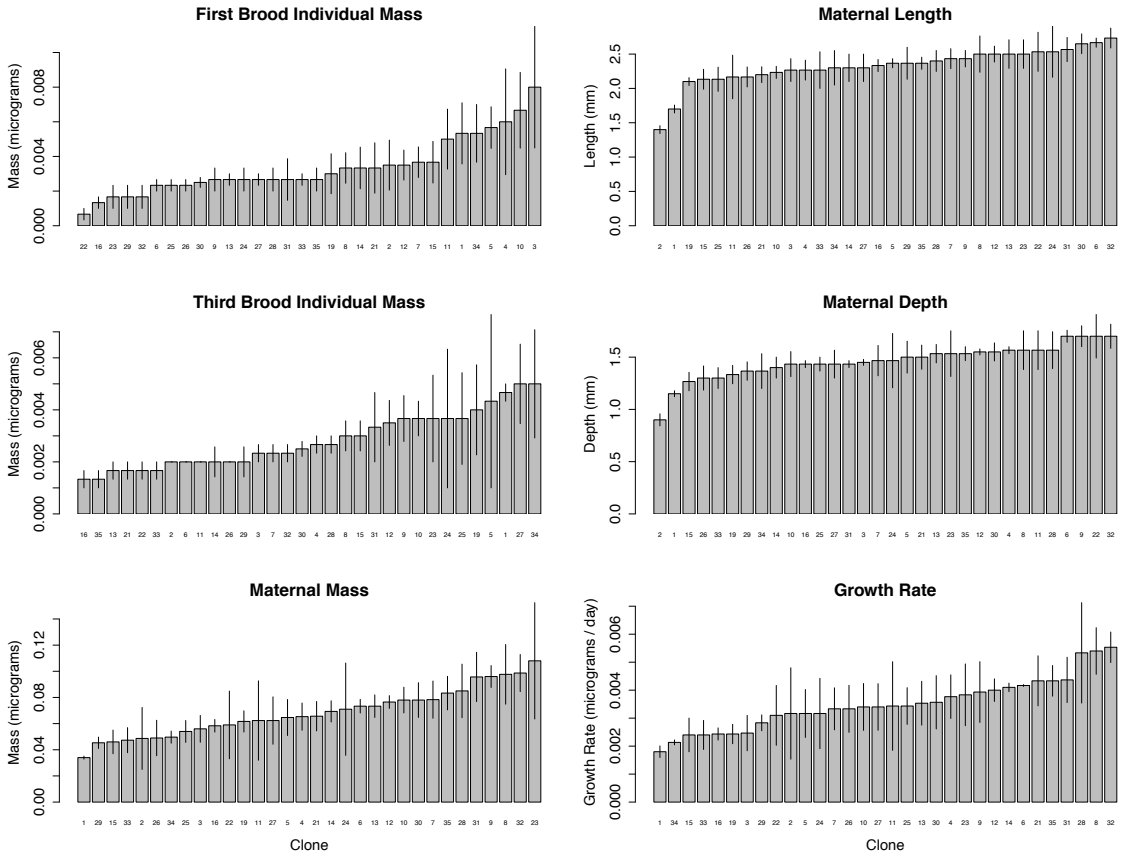


Figure 1.1. Panel variation of *Daphnia* size traits. First and third brood individual mass is the mass of individual offspring, maternal size characteristics are measured at the third brood, and growth rate is calculated from the maternal mass, first brood individual mass, and time from birth to third brood (see Figure 2). Error bars are 95% confidence intervals, and the clone-wise summary data is provided in SI Table 1.1.

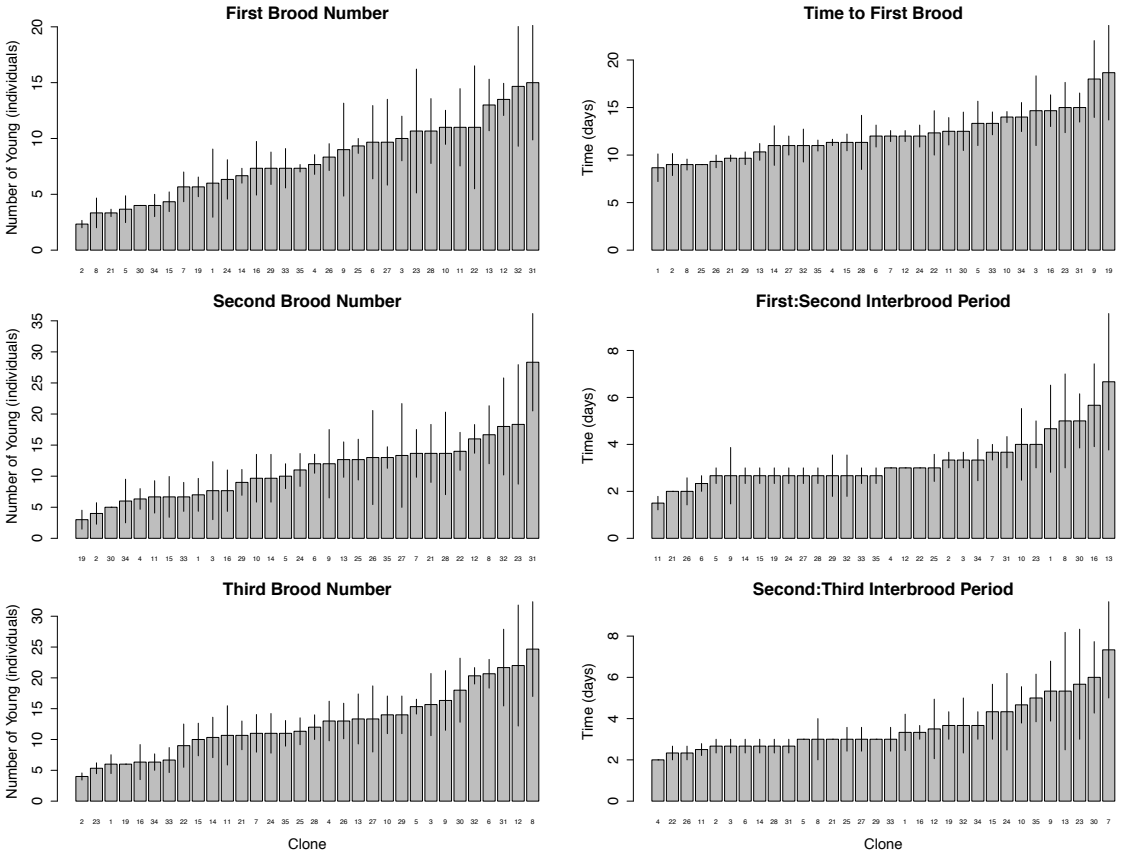


Figure 1.2. Panel variation in the number and timing of offspring. Error bars are 95% confidence intervals, and the clone-wise summary data is provided in SI Table 1.1.

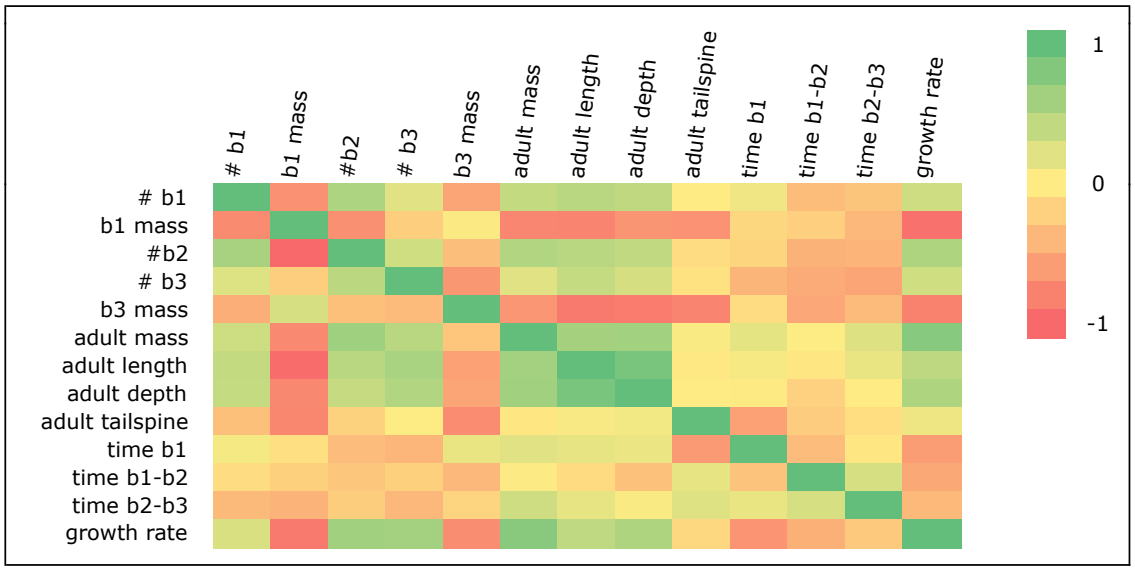


Figure 1.3. Genetic (below diagonal) and phenotypic (above diagonal) inter-trait correlations among growth and fecundity traits. "bN" is the brood number of interest. Note that phenotypic correlations tend to be weaker than genetic correlations, but there are exceptions such as the stronger phenotypic correlation between b3 mass and several other traits.

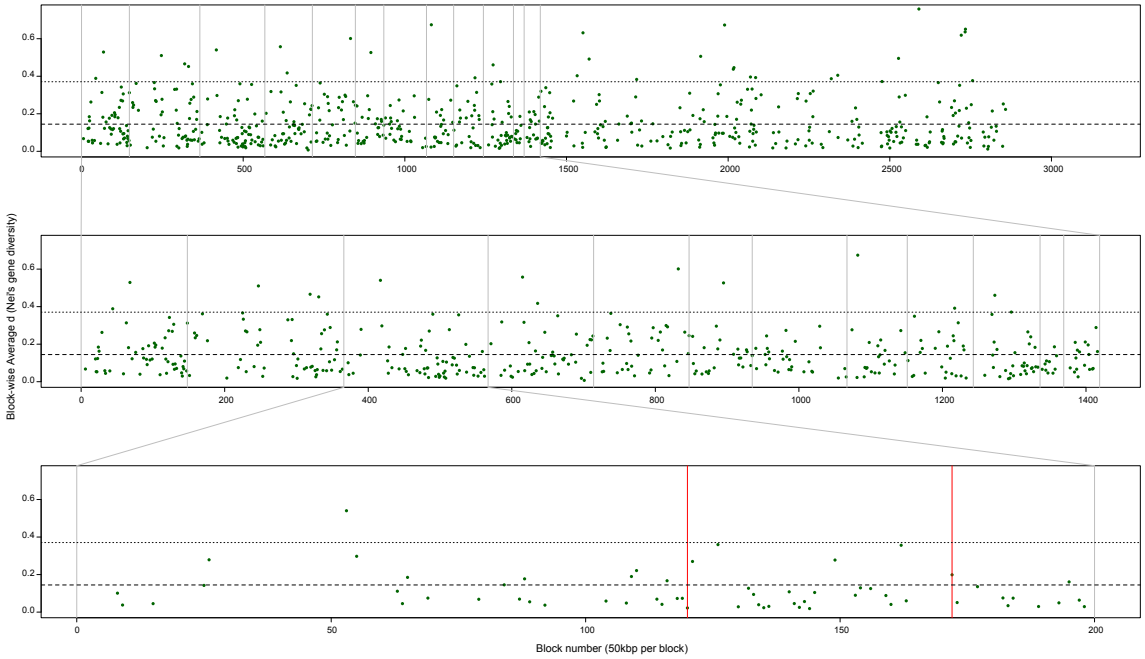


Figure 1.4. Nucleotide diversity (Nei's d) across the genome (A), among scaffolds anchored to linkage groups (B), and within linkage group 3 (C). Light gray vertical lines demarcate linkage group boundaries, and the vertical red lines in panel C mark the window containing LDH genes, which is a common marker for clone identification in ecological studies. Horizontal dashed lines are the mean diversity values and horizontal dotted lines mark the 95th percentile of estimated diversity values. Nucleotide diversity was calculated either in 50-kb non-overlapping windows or over the entirety of a scaffold (if less than 50kb), and only if there were ≥ 3 polymorphic, typed loci in the window/scaffold. Ordering follows the marker order given on the *Daphnia* Genomics Consortium web portal; scaffolds with markers whose order was ambiguous or conflicting relative to the genome assembly were considered unanchored and move to the unanchored section of panel A.

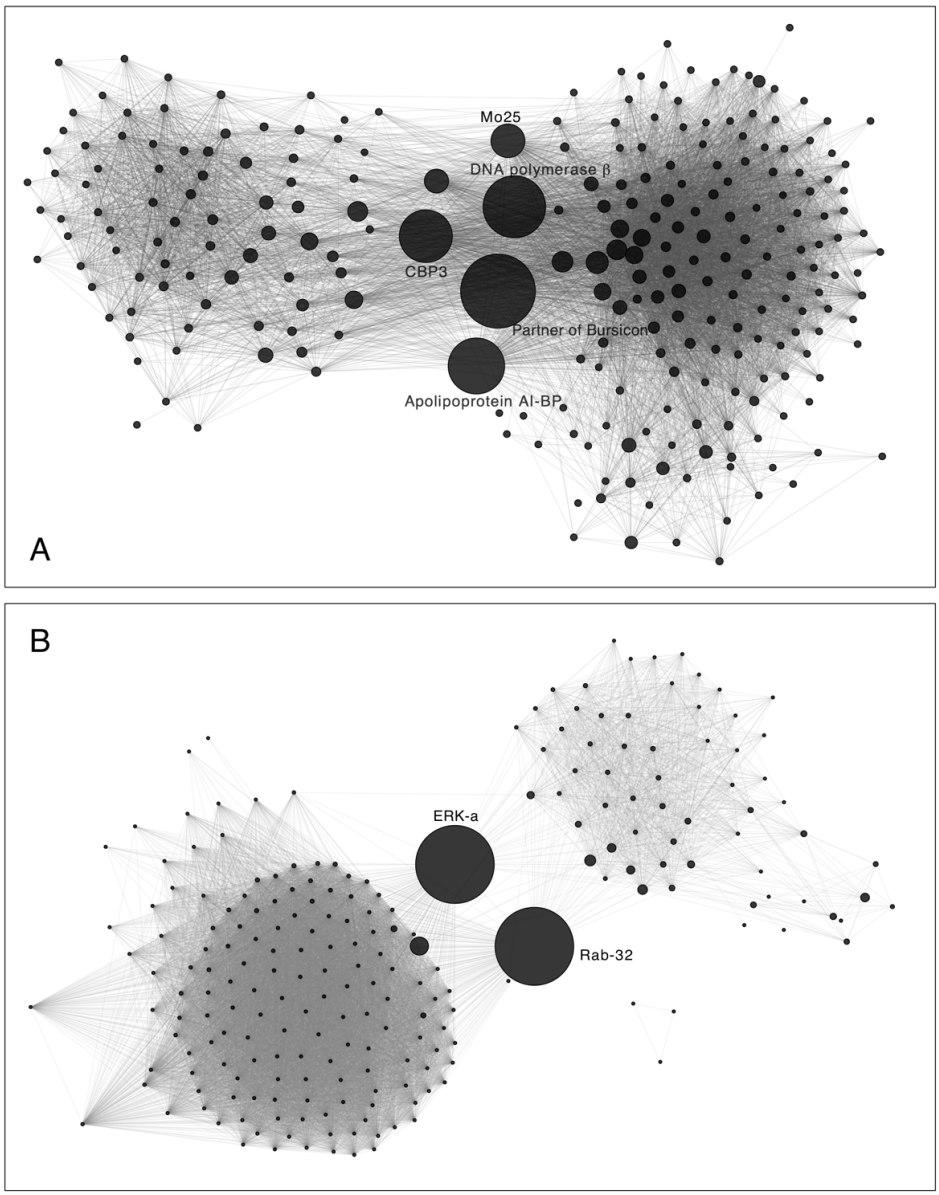


Figure 1.5. Coexpression networks associated with variation in adult mass (A) and third-brood offspring mass (B). Node size is proportional to the betweenness-centrality of the gene; larger nodes connect more groups of coexpressed genes. Select genes with high betweenness-centrality are labeled and discussed in the text.

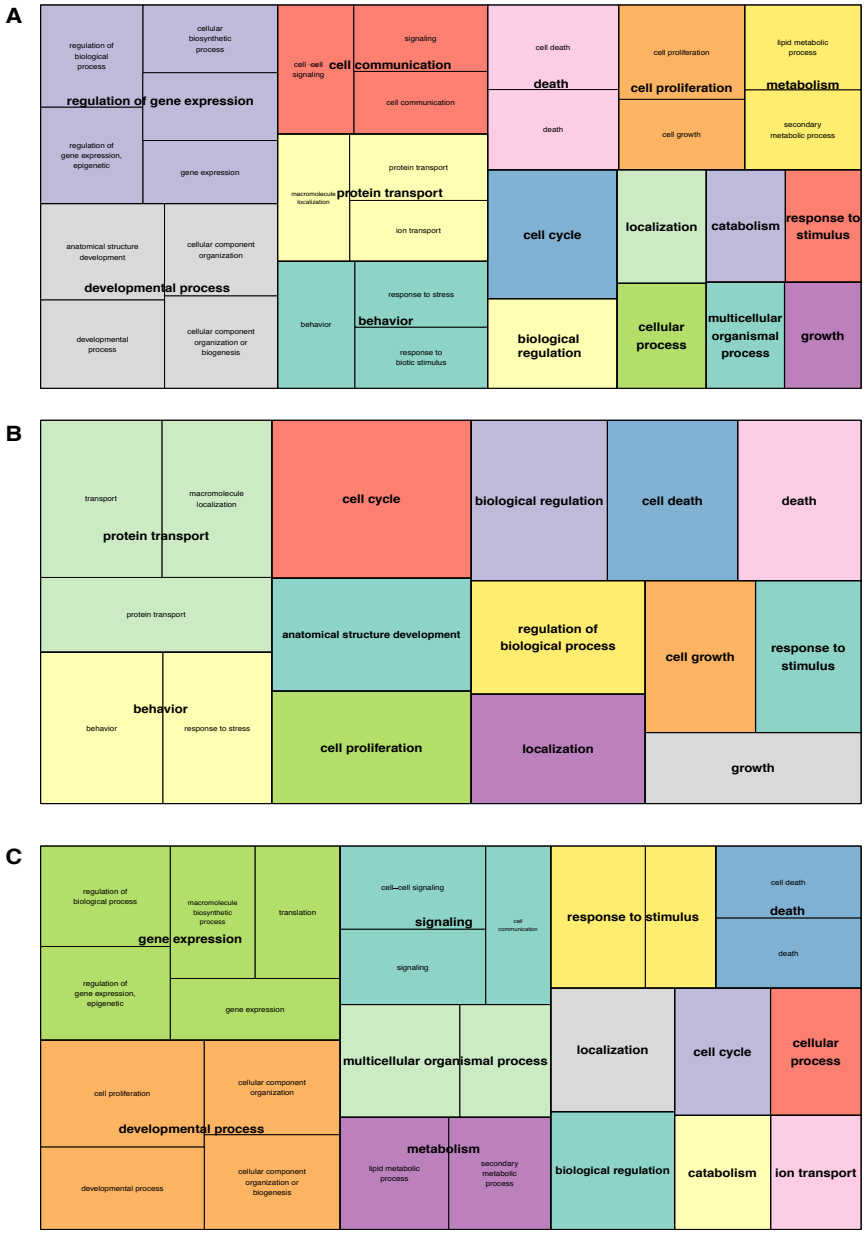


Figure 1.6. Gene Ontology Biological Process term enrichment for all growth and fecundity traits (A), adult *Daphnia* traits (B), and offspring traits (C). Figures are based on the reduction and summary provided by REVIGO (Supek et al. 2011), given the enrichment calculations from GOstats (Gentleman 2011). Large font labels are representative terms for the colored block over which the label is situated, and block size is proportional to the p-value of the enrichment (i.e., larger blocks possess lower p).

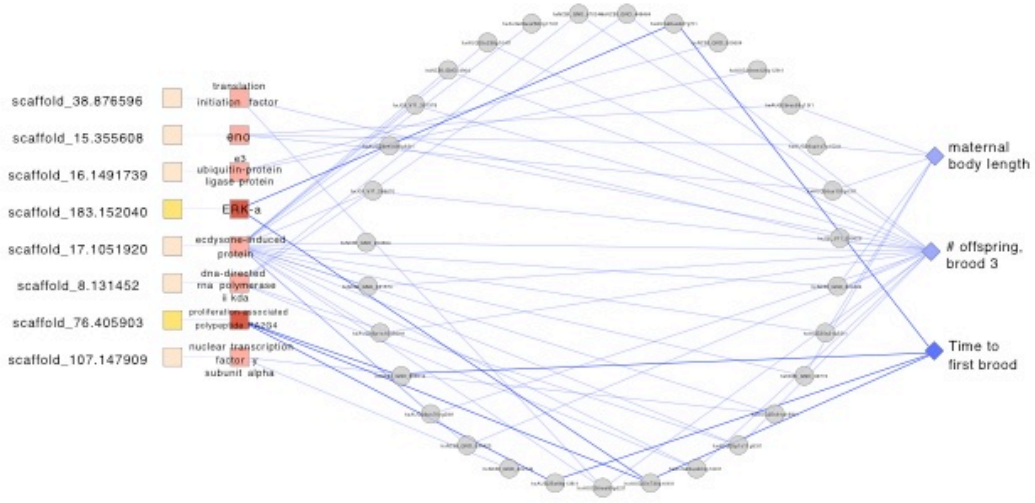


Figure 1.7. An integrated gene network hypothesis for three *Daphnia* growth and fecundity traits. SNPs and the transcriptional regulators in which they are found are represented by squares; differentially expressed genes (DEGs) whose expression is related to marker variation and trait variation are represented by gray circles; and traits are represented by blue diamonds. Edges connecting marker-genes, DEGs, and time to first brood are highlighted darker blue. Labels and edges between coexpressed DEGs are suppressed for figure clarity; the data are available in SI Table X. Highlighted nodes and edges are discussed in the text.

TABLES

Category	Term	Adult size	Offspring size	Offspring number	Interbrood	Growth rate
Metabolism	Carbohydrate metabolism	■	■	■	■	■
	Lipid metabolism	■	■	■	■	■
	Nucleotide/amino acid metabolism	■	■	■	■	■
	Glycan biosynthesis	■	■		■	■
	Secondary metabolite biosynthesis	■	■	■	■	■
Genetic Information Processing	Transcription	■	■	■	■	■
	Translation: Ribosome	■	■	■	■	■
	Folding, sorting, degradation	■	■	■	■	■
Environmental Information Processing	Membrane transport		■	■		
	Signal transduction: mTOR	■	■	■	■	■
	Signal transduction: MAPK	■	■			■
	Signaling molecules: G protein-coupled receptors					
	Signaling molecules: Ion channels					
	Signaling molecules: Cytokine receptors	■				■
Cellular Processes	Transport and catabolism	■		■		■
	Cell growth and death: cell cycle	■	■	■	■	■
	Cell growth and death: apoptosis	■		■		■
Organismal Systems	Immune system: Toll-like receptor signaling	■		■		■
	Endocrine system: Insulin signaling pathway	■	■	■	■	■
	Digestive system: carbohydrate, protein, fat digestion	■	■	■	■	■
	Sensory system: phototransduction					
	Sensory system: taste transduction	■	■	■		■

Table 1.1. Expected trait-pathway relationships for groups of *Daphnia* growth and fecundity traits. Blue cells mark trait-pathway combinations for which I expect to recover a relationship. The categories and terms are derived from KEGG and Gene Ontology.

Trait	H^2	CV_g	V_g	V_p	mean
Growth rate (μg d^{-1})	0.533	3.54E-04	1.24E-06	2.33E-06	0.004
# off. brood 1	0.622	1.989	16.088	25.858	8.089
# off. brood 2	0.566	3.352	37.396	66.112	11.156
# off. brood 3	0.652	2.968	37.554	57.633	12.653
Time to brood 1	0.630	0.683	8.294	13.173	12.146
brood 1: brood 2 time	0.569	0.545	1.773	3.116	3.255
brood 2: brood 3 time	0.548	0.636	2.294	4.184	3.604
brood 1 offspring mass	0.619	0.001	0.000	0.000	0.003
brood 3 offspring mass	0.477	0.001	0.000	0.000	0.003
maternal body depth (mm)	0.534	0.007	0.000	0.001	0.068
maternal body length (mm)	0.675	0.039	0.091	0.135	2.322
maternal mass (μg)	0.640	0.025	0.037	0.058	1.459
maternal tailspine length (mm)	0.765	0.025	0.016	0.021	0.631

Table 1.2. Quantitative genetics of growth and fecundity traits in the *Daphnia* panel. H^2 = broad-sense heritability; CVg = genotypic coefficient of variation; Vg,p = genotypic and phenotypic variances; mean = mean trait value across all replicates of all clones.

Trait	Gene summaries				Modularity summary		
	# genes	# annotated	# unannotated	Pct. annotated	# modules	Min. module size	Max. module size
Growth rate	75	28	47	37.33	4	8	35
# off. brood 1	132	58	74	43.94	4	18	67
# off. brood 2	196	72	124	36.73	3	26	137
# off. brood 3	160	62	98	38.75	5	15	54
Time to brood 1	135	48	87	35.56	6	12	36
brood 1: brood 2 time	99	53	46	53.54	5	10	40
brood 2: brood 3 time	98	43	55	43.88	5	12	30
brood 1 offspring mass	140	54	86	38.57	3	13	110
brood 3 offspring mass	261	111	150	42.53	3	14	185
maternal body depth	152	60	92	39.47	4	21	57
maternal body length	182	63	119	34.62	5	8	85
maternal mass	204	92	112	45.10	5	8	120
maternal tailspine length	198	86	112	43.43	5	9	72

Table 1.3. Summary statistics for differentially expressed genes (DEGs) and coexpression modules associated with variation in the fourteen *Daphnia* growth and fecundity traits. The specific genes and module memberships for each trait are provided in SI Table 5.

LITERATURE CITED

Abrams, P.A. & Rowe, L. (1996). The effects of predation on the age and size of maturity of prey. Evolution, 1052–1061.

Ackerly, D.D. & Cornwell, W.K. (2007). A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. Ecol. Lett., 10, 135–145.

Adamowicz, S.J., Petrusek, A., Colbourne, J.K., Hebert, P.D.N. & Witt, J.D.S. (2009). The scale of divergence: A phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. Mol. Phylogenet. Evol., 50, 423–436.

Alekseev, V. & Lampert, W. (2001). Maternal control of resting-egg production in *Daphnia*. NATURE, 414, 899–901.

Alon, U. (2003). Biological networks: the tinkerer as an engineer. Science, 301, 1866–1867.

Antonovics, J. (2003). Toward community genomics? Ecology, 84, 598–601.

Arendt, J. & Reznick, D. (2008). Convergence and parallelism reconsidered: what have I learned about the genetics of adaptation? Trends Ecol. Evol., 23, 26–32.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet., 25, 25–29.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., et al. (2009). Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet, 41, 299–307.

Barabasi, A.-L. & Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet, 5, 101–113.

Beckerman, A.P., Rodgers, G.M. & Dennis, S.R. (2010). The reaction norm of size and age at maturity under multiple predator risk. J. Anim. Ecol., 79, 1069–1076.

Bernot, R., Dodds, W., Quist, M. & Guy, C. (2006). Temperature and kairomone induced life history plasticity in coexisting *Daphnia*. Aquat. Ecol., 40, 361–372.

Bogdan, K.G. & Gilbert, J.J. (1984). Body size and food size in freshwater zooplankton. Proc. Natl. Acad. Sci., 81, 6427.

Boucher, P., Ditlecadet, D., Dubé, C. & DUFRESNE, F. (2010). Unusual duplication of the insulin-like receptor in the crustacean *Daphnia pulex*. Bmc Evol. Biol., 10, 305.

Bownes, M. (1992). Why is there sequence similarity between insect yolk proteins and vertebrate lipases? J. Lipid Res., 33, 777–790.

Bownes, M., Lineruth, K. & Mauchline, D. (1991). Egg production and fertility in <i>*Drosophila*</i> depend upon the number of yolk-protein gene copies. Mol. Gen. Genet. Mgg, 228, 324–327.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R. & Hafen, E. (2001). An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. Curr. Biol., 11, 213–221.

Brooks, J.L. & Dodson, S.I. (1965). Predation, Body Size, and Composition of Plankton. Science, 150, 28.

Bruggeman, F.J. & Westerhoff, H.V. (2007). The nature of systems biology. Trends Microbiol., 15, 45–50.

Carpenter, S.R. & Kitchell, J.F. (1996). The trophic cascade in lakes. Cambridge University Press.

Cernỳ, M. & Bytel, J. (1991). Density and size distribution of *Daphnia* populations at different fish predation levels. Hydrobiologia, 225, 199–208.

Chase, J.M. & Leibold, M.A. (2003). Ecological Niches: Linking Classical and Contemporary Approaches. University Of Chicago Press.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., et al. (2011). The Ecoresponsive Genome of *Daphnia pulex*. Science, 331, 555 – 561.

Cousyn, C., De Meester, L., Colbourne, J.K., Brendonck, L., Verschuren, D. & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. Proc. Natl. Acad. Sci., 98, 6256.

Cristescu, M.E., Constantin, A., Bock, D.G., Cáceres, C.E. & Crease, T.J. (2012). Speciation with gene flow and the genetics of habitat transitions. Mol. Ecol., 21, 1411–1422.

David, M., Dzamba, M., Lister, D., Ilie, L. & Brudno, M. (2011). SHRiMP2: Sensitive yet Practical Short Read Mapping. Bioinformatics, 27, 1011 –1012.

Dawkins, R. (1982). The Extended Phenotype: The Long Reach of the Gene. Oxford University Press, USA.

Dennis, S.R., Carter, M.J., Hentley, W.T. & Beckerman, A.P. (2011). Phenotypic convergence along a gradient of predation risk. Proc. R. Soc. B Biol. Sci., 278, 1687 –1696.

Duffy, M.A., Hall, S.R., Cáceres, C.E. & Ives, A.R. (2009). Rapid evolution, seasonality, and the termination of parasite epidemics. Ecology, 90, 1441–1448.

Duffy, M.A. & Sivars-Becker, L. (2007). Rapid evolution and ecological host–parasite dynamics. Ecol. Lett., 10, 44–53.

Ellner, S.P., Geber, M.A. & Hairston Jr, N.G. (2011). Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. Ecol. Lett.

Elser, J.J., Acharya, K., Kyle, M., Cotner, J., Makino, W., Markow, T., et al. (2003). Growth rate–stoichiometry couplings in diverse biota. Ecol. Lett., 6, 936–943.

Enquist, B.J., West, G.B., Charnov, E.L. & Brown, J.H. (1999). Allometric scaling of production and life-history variation in vascular plants. Nature, 401, 907–911.

Espinosa-Soto, C. & Wagner, A. (2010). Specialization Can Drive the Evolution of Modularity. Plos Comput Biol, 6, e1000719.

Falcon, S. & Gentleman, R. (2011). How To Use GOstats Testing Gene Lists for GO Term Association.

Feder, M.E. & Mitchell-Olds, T. (2003). Evolutionary and ecological functional genomics. Nat. Rev. Genet., 4, 649–655.

Fisk, D.L., Latta, L.C., Knapp, R.A. & Pfrender, M.E. (2007). Rapid evolution in response to introduced predators I: rates and patterns of morphological and life-history trait divergence. Bmc Evol. Biol., 7, 22.

Fujiwara, M., Sengupta, P. & McIntire, S.L. (2002). Regulation of Body Size and Behavioral State of C. elegans by Sensory Perception and the EGL-4 cGMP-Dependent Protein Kinase. Neuron, 36, 1091–1102.

Gliwicz, Z. & Umana, G. (1994). Cladoceran body-size and vulnerability to copepod predation. Limnol. Ocean., 39, 419–424.

Gliwicz, Z. & Wrzosek, D. (2008). Predation-mediated coexistence of large- and small-bodied *Daphnia* at different food levels. Am. Nat., 172, 358–374.

Goulden, C.E., Henry, L.L. & Tessier, A.J. (1982). Body Size, Energy Reserves, and Competitive Ability in Three Species of Cladocera. Ecology, 63, 1780–1789.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol., 8, R24.

Hairston, N., Lampert, W., Caceres, C., Holtmeier, C., Weider, L., Gaedke, U., et al. (1999). Lake ecosystems - Rapid evolution revealed by dormant eggs. NATURE, 401, 446–446.

Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T. & Fox, J.A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. Ecol. Lett., 8, 1114–1127.

Hall, S., Leibold, M., Lytle, D. & Smith, V. (2004). Stoichiometry and planktonic grazer composition over gradients of light, nutrients, and predation risk. ECOLOGY, 85, 2291–2301.

Hansen, T. (2006). The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst., 37, 123–157.

Hansen, T.F. (2003). Is modularity necessary for evolvability?: Remarks on the relationship between pleiotropy and evolvability. Biosystems, 69, 83–94.

Hart, R.C. & Bychek, E.A. (2010). Body size in freshwater planktonic crustaceans: an overview of extrinsic determinants and modifying influences of biotic interactions. Hydrobiologia, 668, 61–108.

Hartmann, H., Guthöhrlein, E.W., Siebert, M., Luehr, S. & Söding, J. (2013). P-value-based regulatory motif discovery using positional weight matrices. Genome Res., 23, 181–194.

Heckmann, L.-H., Sibly, R.M., Connon, R., Hooper, H.L., Hutchinson, T.H., Maund, S.J., et al. (2008). Systems biology meets stress ecology: linking molecular and organismal stress responses in *Daphnia* magna. Genome Biol, 9, R40.

Heier, C. & Dudycha, J.L. (2009). Ecological speciation in a cyclic parthenogen: Sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia* pulicaria. Limnol. Ocean., 54, 492–502.

Houle, D. (2009). Evolution in Health and Medicine Sackler Colloquium: Numbering the hairs on myheads: The shared challenge and promise of phenomics. Proc. Natl. Acad. Sci.

Houle, D., Govindaraju, D.R. & Omholt, S. (2010). Phenomics: the next challenge. Nat Rev Genet, 11, 855–866.

Ideker, T. & Krogan, N.J. (2012). Differential network biology. Mol Syst Biol, 8.

Ishimoto, H. & Kitamoto, T. (2010). The Steroid Molting Hormone Ecdysone Regulates Sleep in Adult *Drosophila melanogaster*. Genetics, 185, 269–281.

Jean-Christophe, S., Michael, P.E., Ralph, T., Denis, T. & John, C.K. (2011). Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. J. Hered.

Jeyasingh, P.D., Ragavendran, A., Paland, S., Lopez, J.A., Sterner, R.W. & Colbourne, J.K. (2011). How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*. Mol. Ecol., 20, 2341–2352.

Jumbo-Lucioni, P., Ayroles, J., Chambers, M., Jordan, K., Leips, J., Mackay, T., et al. (2010). Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*. Bmc Genomics, 11, 297.

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 28, 27–30.

Kato, Y., Matsuura, T. & Watanabe, H. (2012). Genomic Integration and Germline Transmission of Plasmid Injected into Crustacean *Daphnia* magna Eggs. Plos One, 7, e45318.

Kato, Y., Shiga, Y., Kobayashi, K., Tokishita, S., Yamagata, H., Iguchi, T., et al. (2011). Development of an RNA interference method in the cladoceran crustacean *Daphnia* magna. Dev. Genes Evol.

Klüttgen, B., Dülmer, U., Engels, M. & Ratte, H.. (1994). ADaM, an artificial freshwater for the culture of zooplankton. Water Res., 28, 743–746.

Knight, J.C. (2004). Allele-specific gene expression uncovered. Trends Genet., 20, 113–116.

Kreutzer, C. & Lampert, W. (1999). Exploitative competition in differently sized *Daphnia* species: a mechanistic explanation. Ecology, 80, 2348–2357.

Lampert, W. (2006). *Daphnia*: model herbivore, predator and prey. Pol. J. Ecol., 54, 607–620.

Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

Lee, K.-S., You, K.-H., Choo, J.-K., Han, Y.-M. & Yu, K. (2004). *Drosophila* Short Neuropeptide F Regulates Food Intake and Body Size. J. Biol. Chem., 279, 50781–50789.

Leibold, M.A. (1996). A graphical model of keystone predators in food webs: Trophic regulation of abundance, incidence, and diversity patterns in communities. Am. Nat., 147, 784–812.

Leibold, M. & Tessier, A.J. (1991). Contrasting patterns of body size for *Daphnia* species that segregate by habitat. Oecologia, 86, 342–348.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009a). The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079.

Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., LeProust, E., et al. (2009b). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science, 324, 1210–1213.

Loewe, L. (2009). A framework for evolutionary systems biology. Bmc Syst. Biol., 3, 27.

Loreau, M. (2010). From Populations to Ecosystems: Theoretical Foundations for a New Ecological Synthesis (MPB-46). Princeton University Press.

Lynch, M. (1984). The limits to life history evolution in *Daphnia*. Evolution, 465–482.

Lynch, M. & Walsh, B. (1998). Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA. 980 pp.

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. Nature, 482, 173–178.

Main, T., Dobberfuhl, D. & Elser, J. (1997). N : P stoichiometry and ontogeny of crustacean zooplankton: A test of the growth rate hypothesis. Limnol. Ocean., 42, 1474–1478.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. Trends Ecol. Evol., 21, 178–185.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res., 20, 1297–1303.

Mendive, F.M., Van Loy, T., Claeysen, S., Poels, J., Williamson, M., Hauser, F., et al. (2005). < i> *Drosophila*</i> molting neurohormone bursicon is a heterodimer and the natural agonist of the orphan receptor DLGR2. Febs Lett., 579, 2171–2176.

Messier, J., McGill, B.J. & Lechowicz, M.J. (2010). How do traits vary across ecological scales? A case for trait-based ecology. Ecol. Lett., 9999.

Meyer, E., Aglyamova, G.V. & Matz, M.V. (2011). Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. Mol. Ecol., 20, 3599–3616.

Miner, B.E., Meester, L.D., Pfrender, M.E., Lampert, W. & Hairston, N.G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proc. R. Soc. B Biol. Sci., 279, 1873–1882.

Mitchell, S. & Read, A. (2005). Poor maternal environment enhances offspring disease resistance in an invertebrate. Proc. R. Soc. B-Biol. Sci., 272, 2601–2607.

Nadeau, J.H. & Dudley, A.M. (2011). Systems Genetics. Science, 331, 1015–1016.

Pantel, J.H., Juenger, T.E. & Leibold, M.A. (2011). Environmental gradients structure *Daphnia pulex* × pulicaria clonal distribution. J. Evol. Biol., 24, 723–732.

Peters, R.H. (1986). The ecological implications of body size. Cambridge University Press.

Pfrender, M.E. & Lynch, M. (2000). Quantitative Genetic Variation in *Daphnia*: Temporal Changes in Genetic Architecture. Evolution, 54, 1502–1509.

Roff, D.A. (1992). The evolution of life histories: theory and analysis. Springer.

Rosenblum, E.B., Römpler, H., Schöneberg, T. & Hoekstra, H.E. (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proc. Natl. Acad. Sci., 107, 2113–2117.

Scoville, A.G. & Pfrender, M.E. (2010). Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. Proc. Natl. Acad. Sci., 107, 4260 –4263.

Seger, R. & Krebs, E.G. (1995). The MAPK signaling cascade. Faseb J., 9, 726–735.

Shimizu, Y. & Urabe, J. (2007). Regulation of phosphorus stoichiometry and growth rate of consumers: theoretical and experimental analyses with *Daphnia*. Oecologia, 155, 21–31.

Škunca, N., Altenhoff, A. & Dessimoz, C. (2012). Quality of Computationally Inferred Gene Ontology Annotations. Plos Comput Biol, 8, e1002533.

Spitze, K. (1991). *Chaoborus* Predation and Life-History Evolution in *Daphnia pulex*: Temporal Pattern of Population Diversity, Fitness, and Mean Life History. Evolution, 45, 82–92.

Spitze, K. (1992). PREDATOR-MEDIATED PLASTICITY OF PREY LIFE-HISTORY AND MORPHOLOGY - *CHAOBORUS*-AMERICANUS PREDATION ON *DAPHNIA-PULEX*. Am. Nat., 139, 229–247.

Spitze, K., Burnson, J. & Lynch, M. (1991). The covariance structure of life-history characters in *Daphnia pulex*. Evolution, 45, 1081–1090.

Stearns, S.C. & Magwene, P. (2003). The Naturalist in a World of Genomics. Am. Nat., 161, 171–180.

Sterner, R.W. & Elser, J.J. (2002). Ecological stoichiometry: the biology of elements from molecules to the biosphere. Princeton University Press.

Stocker, H. & Hafen, E. (2000). Genetic control of cell size. Curr. Opin. Genet. Dev., 10, 529–535.

Stuart, J.M., Segal, E., Koller, D. & Kim, S.K. (2003). A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. Science, 302, 249–255.

Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. Plos One, 6, e21800.

Székely, G.J. & Rizzo, M.L. (2009). Brownian distance covariance. Ann. Appl. Stat., 1236–1265.

Székely, G.J., Rizzo, M.L. & Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. Ann. Stat., 35, 2769–2794.

Taub, M.A., Corrada Bravo, H. & Irizarry, R.A. (2010). Overcoming bias and systematic errors in next generation sequencing data. Genome Med, 2, 87.

Tessier, A. & Goulden, C. (1982). Estimating food limitation in cladoceran populations. Limnol. Ocean., 27, 707–717.

Tessier, A., Young, A. & Leibold, M. (1992). Population dynamics and body-size selection in *Daphnia*. Limnol. Ocean., 37, 1–13.

Thompson, J.N. (1998). Rapid evolution as an ecological process. Trends Ecol. Evol., 13, 329–332.

Tollrian, R. (1993). Neckteeth formation in *Daphnia pulex* as an example of continuous phenotypic plasticity: morphological effects of *Chaoborus* kairomone concentration and their quantification. J. Plankton Res., 15, 1309–1318.

Tollrian, R. & Leese, F. (2010). Ecological genomics: steps towards unraveling the genetic basis of inducible defenses in *Daphnia*. Bmc Biol., 8, 51.

Turner, T.L., Stewart, A.D., Fields, A.T., Rice, W.R. & Tarone, A.M. (2011). Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. Plos Genet., 7, e1001336.

Urabe, J., Kyle, M., Makino, W., Yoshida, T., Andersen, T. & Elser, J. (2002). Reduced light increases herbivore production due to stoichiometric effects of light/nutrient balance. Ecology, 83, 619–627.

Wang, S., Meyer, E., McKay, J.K. & Matz, M.V. (2012). 2b-rad: a simple and flexible method for genome-wide genotyping. Nat. Methods, 9, 808–810.

Webb, C.T., Hoeting, J.A., Ames, G.M., Pyne, M.I. & LeRoy Poff, N. (2010). A structured and dynamic framework to advance traits-based theory and prediction in ecology. Ecol. Lett., 13, 267–283.

Weider, L.J., Makino, W., Acharya, K., Glenn, K.L., Kyle, M., Urabe, J., et al. (2005). Genotype X environment interactions, stoichiometric food quality effects, and clonal coexistence in *Daphnia pulex*. Oecologia, 143, 537–547.

West, G.B., Brown, J.H. & Enquist, B.J. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organisms. Science, 284, 1677–1679.

Wetterer, J.K. (1989). Mechanisms of prey choice by planktivorous fish: perceptual constraints and rules of thumb. Anim. Behav., 37, 955–967.

Wheat, C. (2010). Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. GENETICA, 138, 433–451.

Wolinska, J., Löffler, A. & Spaak, P. (2007). Taxon-specific reaction norms to predator cues in a hybrid *Daphnia* complex. Freshw. Biol., 52, 1198–1209.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., et al. (2003). The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol., 20, 1377–1419.

Wright, D.I. & O'Brien, W.J. (1982). Differential location of *Chaoborus* larvae and *Daphnia* by fish: the importance of motion and visible size. Am. Midl. Nat., 68–73.

Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F. & Hairston, N.G. (2003). Rapid evolution drives ecological dynamics in a predator-prey system. Nature, 424, 303–306.

Chapter 2: Exploration of the molecular basis of behavioral variation in the model ecological species, *Daphnia pulex*

Variation in behavioral traits is central to mediating many interactions between individuals and their environment, which may include the influences of abiotic factors and both con- and heterospecifics. Several swimming-related traits are important determinants of the niche of the waterflea (*Daphnia pulex*), a classical model ecological species, and variation in these traits influences landscape-scale distribution of different genotypes. Using a diverse panel of wild-caught *Daphnia* clones, I linked variation in 18 behavioral traits to genotype and gene expression variation. These traits vary by as much as 6-fold in the panel and tend to possess high broad-sense heritability (0.48-0.80). Several a priori genes and pathways expected to be associated with behavioral variation, including globins and protein transport-associated genes, are among 2246 candidate genes (from 6434 differentially-expressed genes) and 180 SNPs that are strongly-correlated with the phenotypic variation. I present putative gene coexpression networks underlying trait variation, as well as a candidate systems genetic network relating sequence variation to expression and organismal phenotypic variation. These results are the first of their kind for this well-established ecological model organism, and emerging model genomic species, and establish a foundation for deeper investigation of the causes of ecologically important behavioral variation.

INTRODUCTION

Behavioral variation is a central aspect of animals that affects ecologically and evolutionarily important interactions with con- and heterospecifics. The heritable component of that variation is subject to the standard evolutionary forces (Barton et al. 2007; Hamilton 2009), and determining the loci, genes, and networks underlying behavioral variation is a major goal for both basic (Jordan et al. 2007; Renn et al. 2008; Simon et al. 2011; Christie & McCoole 2012) and applied research (McGuffin et al. 2001; Inoue & Lupski 2003; Robinson et al. 2008; Bell 2009). Furthermore, investigating the molecular basis of behavioral variation in the context of a species' ecology is central to understanding the factors shaping—and shaped by—heritable variation.

The common waterflea, *Daphnia pulex* (group), is a small crustacean with widespread distribution in Northern Hemisphere freshwater ponds and lakes. For at least a century (Weismann 1876 in Lampert 2006; Forbes 1880) ecologists have used the species group as a model for competition (Dodson 1974; Cooper & Smith 1982; DeMott & Kerfoot 1982; Leibold 1991; Duffy 2010), predation (Dodson 1974; Leibold 1991; Leibold & Tessier 1991; Tessier et al. 1992; Gliwicz & Umana 1994; Mikulski & Pijanowska 2010), ecological stoichiometry (Main et al. 1997; Acharya et al. 2004; Hall et al. 2004; Jeyasingh et al. 2011; Miner et al. 2012), and ecotoxicological studies (Enserink et al. 1991; Rouvalis et al. 2009; McCoole et al. 2011; Christie & McCoole 2012). Importantly, the relationship between *Daphnia* trait variation and ecological variation is reasonably well-understood, and ranges from the effects of size and shape on competitive ability and predator susceptibility (Dodson 1974; Cooper & Smith 1982; Lampert 1993, 2006; Gliwicz & Umana 1994) to the various roles of phenotypic

plasticity (Spitze 1992; Lampert 1993; Boersma et al. 1998; Scheiner & Berrigan 1998; Hülsmann et al. 2011; Miner et al. 2012), and, central to this research, swimming behaviors (Figure 2.1).

At the large scale (relative to *Daphnia* size), the behavior of diel vertical migration (DVM) in the water column has long been known to affect coexistence with vertebrate predators. Non-migratory epilimnetic genotypes are susceptible to predation by fish, whereas migratory genotypes reside in the hypolimnion, under reduced predation pressure, during the day (references for these and subsequent statements are provided in Table 1). DVM consists of two parts: the "constitutive" behavior driven by phototaxis, and the behavior as altered by chemical cues from fish (generically termed "kairomones"). In heterogeneous landscapes containing both fishless and fish-containing ponds, quantifying DVM under both conditions is necessary for determining the probability that a particular genotype can persist at a location. In addition to its role in predator avoidance, DVM strongly affects the stoichiometric quality of algae availability (Hood & Sterner 2010). Lower-quality food reduces *Daphnia* growth rates and size, which, at least under abundant food conditions, reduces competitive ability (Tessier & Goulden 1982).

Large-scale diel migration phenomena are better-studied than the effects of small-scale swimming behavioral variation, but this latter topic has also been investigated (Table 1, bottom half). Because *Daphnia* are filter feeders, swimming speed strongly influences the rate at which algae (and other prey) are consumed, and alters competitive abilities. Higher swimming speeds should increase the encounter rate with pathogens (Hall et al. 2007), although *D. pulex* tends not to be as susceptible to parasites as *D. dentifera*. Most other small-scale swimming behaviors have been linked to variation in

predation susceptibility, e.g., fast and erratic swimmers are more likely to be noticed by fish predators, but are more likely to escape predation attempts by invertebrate predators (Drenner et al. 1978; Pijanowska et al. 1993). In addition to steady-state swimming behavior, the reactivity to predation strikes is critically important to predation susceptibility. The reaction traits in the context of gene expression are necessarily important from the standpoint of constitutive gene expression studies: if the appropriate reaction (e.g., fast escape) is not realized in a matter of one to a few seconds, there is no chance for transcriptional changes that can alter the behavior before an individual is dead.

The *D. pulex* group is phenotypically and ecologically diverse, and has traditionally been considered two species and a hybrid: *D. pulex* characteristically occupies shaded, fishless ponds; *D. pulicaria* characteristically occupies deep, fish-laden lakes; and hybrids between the two occupy sunny, fishless ponds. Throughout I use *Daphnia* to generically refer to clones of all three ecotypes given recent phylogenetic data (Adamowicz et al. 2009), breeding and multilocus analyses (Heier & Dudycha 2009; Cristescu et al. 2012), and my recovery of both low population structure and continuous trait variation across clones of all ecotypes in the present work.

Expectations

While genome-wide data collection is a prominent tool for developing novel hypotheses about how molecular variation affects phenotypes, the inferences from such approaches are strengthened when there are a priori expectations. Here I review some of the gene families I anticipate may be related to *Daphnia* swimming behavior variation, and provide a summary of expected molecular pathway – trait relationships in Table 2.

Macro-scale swimming behavior such as DVM, outside the chemical cues of predators, is thought to be driven primarily by variation in incident light. The *Daphnia* genome is characterized by expansion of gene families encoding photoreactive proteins, including cryptochromes, opsins, and G proteins (Colbourne et al. 2011). Because cryptochtomes are blue-light sensitive and the blue portion of the visible spectrum has the highest transmission rate in the water column, we expect cryptochrome variation to play an important role in shaping DVM variation (Tilden et al. 2011). Cryptochromes are central to the circadian clock, and between-clone variation in their activity may systematically affect the expression (and evolution) of other circadian genes such as CLOCK, CYCLE, and PERIOD. Even outside of experimental manipulations of light levels/cycles and concomitant RNA collection (see Methods), we may expect clones that vary in phototactic behavior to vary in the expression of these gene families.

Macro-scale swimming behavior is also influenced by the presence of the chemical cues of predators. Genes related to chemoreception, in particular gustatory receptors (Grs) in *Daphnia* (Penalva-Arana et al. 2009), are the first key in the transcriptional cascade that regulates response. A recent review of taste pathways indicates that it is unknown whether Grs are active through G protein coupled receptors or as ion channels (Yarmolinsky et al. 2009); coexpression of Grs and GPCRs in *Daphnia* may provide a clue.

In addition to the proximate signaling pathways expected to drive variation in movement, we anticipate that clones must be physiologically prepared for systematic variation in the limnetic environment, including hypoxic conditions in the epilimnion (Wetzel 2001); I expect, then, that globins such as hemoglobin (Duffy 2010) and

cytoglobin may be differentially regulated among clones in proportion to their DVM and kairomone responses.

Micro-scale swimming behavior and reaction to tactile stimulus may be most easily relatable to general movement and "locomotor reactivity" in *Drosophila* and other model eukaryotes. The classical behavioral genetics gene foraging, which is cGMP-dependent kinase in *Drosophila* and which shows high sequence similarity to a pair of cAMP-dependent kinases in *Daphnia*, expresses distinctive phenotypes of rarely- or rapidly-moving fly larvae and adults (Pereira & Sokolowski 1993). I anticipate a correlation between expression of these (or perhaps other) cAMP-dependent kinases and steady-state swimming speed and patterns in *Daphnia*. Ayroles et al. (2009) found GO enrichment for locomotor reactivity in several Biological Processes in *Drosophila melanogaster* (top hits include, e.g., lipid metabolism and physiological process); Cellular Components (membrane, integral to membrane, and intrinsic to membrane); Molecular Function (oxidoreductase activity and transporter activity); and GO keywords of "transport", "ion transport", and "oxidoreductase". These ontologies are logical given the need to respond to mechanical stimulus, and I expect similar enrichments for genes and loci whose expression and genotypic variation are correlated with *Daphnia* response to simulated attack.

Present goal

My goal in this work is to provide a first-look at the genetic and gene expression basis of swimming behavior variation in *D. pulex*. The species' genome was recently sequenced (Colbourne et al. 2011) and there are exceptionally few genetic resources available to inform my understanding of the molecular basis of heritable trait variation in

Daphnia; exploratory genomic approaches are ideal for discovering new associations and generating focused hypotheses (Ayroles et al. 2009; Stone & Ayroles 2009). To this end, I employed a diverse panel of 32 *D. pulex* (complex) clones from southern Michigan, USA, and measured swimming trait, genetic, and expression variation under common garden conditions. This panel is also the focus of similar work with life history traits (size, growth, and fecundity) reported in Chapter 1. I find support for many expected pathway-trait relationships; describe candidate coexpression networks and genetically variable loci; and identify candidate genes both with and without functional annotation transferred from classical model genetic species. I close with a discussion of the implications of the results for our understanding of *Daphnia* ecology.

RESULTS

Phenotypic variation

The panel of *Daphnia* clones consists of two nominal *D. pulicaria*, fourteen nominal *D. pulex*, and 16 hybrids based on collection locations. These varied more than four-fold in their constitutive phototaxis, phototaxis when exposed to fish kairomones, and in the degree of phototaxis plasticity (Figure 2). Some clones identified as "hybrids" (e.g., Clone 8) possessed stronger negative phototaxis in response to fish kairomones than nominal *D. pulicaria* (e.g., Clone 2), whereas some clones exhibited rather maladaptive response to fish kairomones, swimming higher in the water column when kairomones were present that when absent (e.g., Clone 3).

The micro-scale swimming behaviors of the panel exhibited 1.6- to 10-fold variation (Figure 3; SI Table 1). While most clones reacted to the attacks by increasing

their mean swimming speed, several decreased their maximum swimming speed post-attack; this is reflected in the fact that over 1/3 of clones decreased the variability (CV) of swimming speed after the attack. In addition, while there was six-fold variability in sinuosity pre-attack, most clones began swimming much more directly (i.e., decreased their sinuosity) after the attack, which resulted in a larger area occupied after the attack.

Broad-sense heritability ranged from 0.48-0.80 for the swimming behaviors (Table 3). Phenotypic and genetic correlations among the traits was modular within trait class, and ranged from -0.63 to 0.94 (Figure 4). I further explore patterns of the shared variation (i.e., principal components) in forthcoming papers.

Genome-wide genetic and gene expression variation

We genotyped the panel at ~5.5 million base positions across all clones, of which 665678 loci were typed in >75% of clones and 6338 were polymorphic in the panel. Heterozygosity was low, at 0.14% on average across clones (range = 0.07-0.2%). Viewed in 50kb non-overlapping windows, there were few genomic regions exhibiting extensive Hardy-Weinberg disequilibrium or other patterns of nucleotide diversity (SI Figure 2.1). Although the number of typed loci at a given distance was relatively low (< 2), a moving average analysis suggests LD decay occurs in ca. 200bp (SI Figure 2.2). Principal components decomposition of the genotype matrix (~5700 polymorphic loci typed in >75% of clones) and STRUCTURE analysis revealed relatively little population structure in the panel (SI Text 2.1). A methodological implication of the weak structure is that genome-wide association analyses are tractable within the limits of the panel size and lack of controlled crosses.

We obtained an average of 1.4 million RNA-seq reads per replicate after filtering, which mapped to > 30000 genes. Of these, ~15000 possessed an average expression level > 5 reads per million mapped and were retained for further analysis. Of the ~15000 genes, 6434 were differentially expressed (DE) at $p < 0.05$ and Benjamini-Hochberg false discover rate control at 1%. I used the WGCNA package for R (Langfelder & Horvath 2008), which clusters genes so that within-cluster correlations are stronger than between-module correlations, to identify coexpression modules among the DEGs. These DE genes (DEGs) clustered into 24-27 global coexpression modules across a wide variety of parameter options, and subsequent analyses are based on the 27-module solution presented in Chapter 1 (SI Figure 1.2). The range of differential expression per gene and among clones was estimated to be 1.3- to 7000-fold, due only to genetic differences between clones.

DEGs and networks associated with swimming behavior

We identified 2246 genes whose expression was strongly distance correlated (Székely et al. 2007) with variation in ≥ 1 behavioral trait. After removing likely spurious relationships there were 71-404 genes per trait, which clustered into 1-12 trait-specific coexpression modules (Table 4, SI Table 2.2). Although one of *Daphnia*'s 46 opsins (hxAUG26rep1s6g338t1, a Rhabdomeric LongB opsin; Colbourne et al. 2011 Table S32) was differentially expressed, it did not correlate with phototaxis variation. Five gustatory receptor genes were differentially expressed; three were associated with variation in plasticity of area occupied while swimming and the area occupied after attack, but were not associated with phototaxis under kairomone exposure. Eight differentially expressed globin genes, including cytoglobins, hemoglobins, and globins,

were strongly correlated with both micro-scale speed-associated swimming behaviors and phototaxis (constitutive and phototaxis plasticity).

The coexpression modules associated with the behavior traits contained genes drawn from all 27 of the global coexpression modules, but only 23 modules where at least ten genes, or 10% of genes, were associated with a trait (SI Table 2.3). Of particular interest is that the "blue", "black", "salmon", and "orange" global coexpression modules are particularly over-represented among the swimming behavior traits relative to life history traits (see Chapter 1).

The expression architecture of trait-specific coexpression networks varied substantially among traits, and even within trait groups. For example, while the number of genes associated with constitutive phototaxis was substantially smaller than the number of genes associated with kairomone exposure response, the network is spread among nine modules and possesses several "hub" genes with high betweenness centrality (a measure of the importance of a node in connecting other nodes across a network; Figure 2.5A). In contrast, the more gene-rich kairomone phototaxis coexpression network is condensed into four modules, one much larger and more-densely connected than the others; furthermore, there is a single gene (an ncRNA, hxAUG25p1s10g224t1) with very high betweenness centrality connecting the large, dense module to a second module (Figure 2.5B). Similarly, there are fewer genes associated with constitutive maximum swimming speed compared to post-attack maximum swimming speed; in contrast to the pattern of modularity of phototaxis genes, these fewer genes group into fewer clusters than for post-attack max speed. The modularity, link density (average number of edges per node), and distribution of "hub" genes is relatively more-even for

the larger post-attack network (Figure 2.6A) than for the smaller constitutive swimming speed network (Figure 2.6B).

While most DEGs associated with each trait are un-annotated (52-72%, $\bar{x}$ = 61.2%), I identified 28 differentially expressed transcription factors whose expression was tightly correlated with trait variation and which constitute candidates for regulating the coexpression of trait-specific modules. Some of these regulators, such as FOXL2, are significantly associated with several micro-scale swimming traits but not macro-scale swimming, suggesting high specificity and strong candidates for driving variation. Other TFs, such as the *Daphnia* ortholog of tailless, are associated with several micro- and macro-scale behaviors, suggesting more general regulatory action.

Given the sets of DEGs associated with variation in each trait, I identified enrichment for 6-78 Gene Ontology terms associated with each trait (SI Table 2.4). Across all 18 traits there was enrichment for Biological Process terms including those related to protein modification process (e.g., "gene expression", "DNA metabolic process"); protein transport; response to stress; and cell communication, among others (Figure 2.7A). Genes associated with macro-scale swimming behavior—phototaxis—were enriched for numerous "protein modification process"-like terms; response to stress; and secondary metabolism, among others (Figure 2.7B). The largest groups of similar GO terms for the micro-scale swimming behaviors include translation-associated terms; response to stress; and protein transport, with smaller representation for genes associated with response to stimulus and cell communication (Figure 2.7C). REVIGO-based treemaps of term enrichments for the Molecular Function and Cellular Component GO domains, as well as for trait-specific enrichments, are provided in SI Figure 2.3.

One possible cause of coexpression is shared expression regulators, and by extension, I anticipate that coexpressed genes share regulatory motifs in promoter regions. I identified between two and 21 promoter motifs based on sequence similarity per swimming trait, with one to 13 motifs associated with each trait-specific coexpression module (SI Table 2.5), and a strong correlation between number of genes per trait and number of motifs ($r = 0.67$, $p = 0.0018$). Some motifs appear multiple times, and are associated with multiple traits. For example, several motifs similar to the C-rich motif associated with the "turquoise" coexpression module of constitutive swimming speed CV, YCCCCCCCCCCCCWV (E-value = 1.88e9), are found in modules for post-attack area occupied and swimming speed CV, plus mean swimming speed and area occupied plasticity. I used TOMTOM (Gupta et al. 2007) to determine if my computationally identified motifs possessed high sequence similarity to motifs known from chromatin immunoprecipitation experiments in *D. melanogaster*. An apparent orthologous sequence of this motif in *Drosophila* (E-value = $5.325e^{-7}$) is the target of the un-named transcriptional regulator gene CG7368; the *Daphnia* ortholog of CG7368 (hxAUG26us24g213t1) is differentially expressed and strongly associated with a reproduction trait, but falls just below statistical significance for behavior-related traits ($dcors < 0.37, p > 0.05$).

Genetic variants associated with swimming behaviors and DEGs

We identified 1-18 markers associated with variation in each of ten traits using a genome-wide association (GWA) analysis of 3450 SNPs with a minor allele frequency ≥ 10%, of which only three markers associated with trait variation fall outside of open reading frames (SI Table 2.6). The 18 markers associated with post-attack maximum

swimming speed were spread across 15 scaffolds; all but two were separated by > 100kb. The exception was two markers on scaffold 10 (at 880459 and 884591 bp) that are located in translation-related genes, MRPS14 (a mitochondrial ribosomal protein, S14) and EEF1A1 (a translation elongation factor), respectively. I identified 4787 marker-DEG pairs for which allelic variants strongly (nominal $p < 1e^{-7}$) predicted expression; of these, 4125 marker-DEG pairs included swimming behavior-associated genes.

We combined the marker-trait and marker-DEG results to derive a first-pass working hypothesis of the systems genetics of *Daphnia* swimming behavior (SI Table 2.7). Although any of the statistically supported markers are candidates, for visualization I focused on the thirteen markers within transcription factors or genes known to regulate transcription, and that were strongly related to genes whose expression was associated with swimming behavior variation (Figure 2.8). This supports a many-marker to many-DEG, and many-DEG to many-trait mapping from genotype to organismal phenotype—which is expected for complex behavioral traits—but more importantly, is a set of hypotheses to be tested as *Daphnia* develops as a model genomic and genetic system.

DISCUSSION

Behavioral variation is a central aspect of ecology, evolution, and human health, and identifying the genes underlying the heritable portion of this variation is a broad goal of biological research. I used a classical model ecological species, *Daphnia pulex* (group), and its newly developed genomic resources to investigate the genes underlying variation in 18 behavioral traits known or believed to play significant roles its niche and landscape-scale distribution. I identified > 2200 genes whose expression is associated with variation in 18 behavioral traits, and recovered support for several a priori candidate

genes and pathways. The high estimated heritabilities of the traits is promising for further developing my understanding the genetics of behavioral variation in *Daphnia*, and variation of between-trait genetic correlations suggests *Daphnia* as a good model for investigating the genetics of pleiotropy.

The estimated linkage disequilibrium decay rate in my *Daphnia* panel falls between the rate for strictly outcrossing *Drosophila* (10bp; Mackay et al.2012) and the largely selfing Arabidopsis (10kb; Nordborg et al. 2002). Perhaps more interesting, and worthy of noting outside the implications for behavioral variation, is that I found relatively little genetic structure in the panel. As noted in the Introduction, the *D. pulex* group has long been considered by ecologists to consist of two parental species (with *D. pulicaria*) and hybrids, each characteristic of distinct habitats and each identified by the genotype at the Ldh locus (Hebert et al. 1988). Recent breeding (Heier & Dudycha 2009) and genetic work (Cristescu et al. 2012) provide a mechanism for gene flow between the putative parental species and molecular support for gene flow despite ecological differentiation, respectively. The weak structure among ~5000 markers supports the other recent research: there are more alleles shared among clone ecotypes than private to each, suggesting there is still gene flow across the ecotypes. The main caveat to this interpretation is that the panel size and the geographic range across which the clones were collected is relatively small. Future population genomic work with more samples collected from across a larger area will help resolve the extent of gene flow in the *D. pulex* group.

With 70% of *Daphnia* gene models currently lacking functional annotation and no clear orthology to genes known in other model organisms, I anticipated and succeeded in uncovering many novel candidate genes related to behavioral variation. However, I also

recovered several expected gene-trait or pathway-trait relationships: a role for globin activity among clones that occupy waters with a range of hypoxic conditions; protein transport-associated Gene Ontology terms; and membrane-associated Cellular Component terms. The lack of evidence for opsin and gustatory receptor gene expression associated with phototaxis and kairomone response suggests that between-clone differences in these phenotypes may be based on protein-coding changes rather than changes in the level at which the genes are transcribed (see also Cristescu et al. (2012)). An alternative hypothesis is that the common-garden nature of the expression assays missed changes in induced expression that are central to behavioral variation; testing this hypothesis will require additional experiments with a smaller number of clones in more environments.

The advent of genome-wide assays has brought with it the persistent challenge of sorting through and interpreting large lists of DEGs associated with variation at other levels of biological organization. Making biological sense of these voluminous results requires simplification, often using Gene Ontology (and related ontologies) term enrichment (Rhee et al. 2008). A posteriori examination of trait-specific and trait-group term enrichment reveals relatively few surprises. As a whole, genes with functions for metabolic processes (e.g., protein, macromolecule, and nucleic acid), response to stress and stimulus, and cellular communication terms are over-represented among genes associated with behavioral variation. These Biological Process GO terms are complemented by Molecular Function terms including kinase activity, signal transduction, peptidase activity, and macromolecule (e.g., protein, carbohydrate, and lipid) binding. All of these make "sense" given the metabolic requirements of swimming at large and small scales, and the need to sense external stressors—the activity of

predators and the availability of oxygen in different limnetic zones—that are central to *Daphnia* ecology (see Table 1). Trait-specific enrichments are similarly justifiable: for example, the roles of protein transport, macromolecule metabolism, and response to stimuli with respect to phototaxis plasticity. The caveat to my interpretations is that most genes associated with behavioral variation do not possess any functional annotation. Identifying the molecular functions of these un-annotated genes will greatly expand my understanding of the molecular basis of swimming variation in *Daphnia*, and the molecular evolution of genes shaping behavioral variation across taxa.

Recent debates about the "loci of evolution" highlight the need to identify not just the presence of variation in intermediate phenotypes—such as gene expression—but also the genetic basis of that variation. Hoekstra and Coyne (2007), for example, advocate that the loci of evolution are predominantly in protein-coding sequence, while "evo-devo" proponents highlight cis-regulatory changes (Stern & Orgogozo 2008, 2009; Carroll 2009). my data are not sufficient to directly address this issue, but the identification of shared promoter motifs and intragenic variants associated with expression and trait variation provide new hypotheses that can be tested identify the loci of evolution in the ecologically well-defined *Daphnia*.

Given the weak population structure of the panel I performed a limited GWA analysis and identified several loci whose variants were predictive of expression and trait variation. Given the small panel size and lack of controlled crosses I emphasize the discovered markers as candidates that warrant further investigation, not the causal loci. Among the markers falling within genes, and associated both with swimming variation and variation in other genes whose expression was correlated with swimming variation, three examples stand out. First, a SNP at scaffold 3:1715566 falls within the gene

encoding protein *slowmo*, and is predictive of expression level of twelve DEGs that are significantly associated with *Daphnia* post-attack swimming speed and area occupied (Figure 2.9A). *slowmo* has been shown experimentally to affect larval *Drosophila* locomotor behavior and speed (Carhan et al. 2004). Second, a SNP at scaffold 87:168746 is in *nonmuscle myosin essential light chain*, and the two alleles are predictive of variation in 85 DEGs strongly correlated with pre-, post-, and change in swimming speed (and area) behaviors (Figure 2.9B). Myosins are ATP-driven motor proteins central to muscle contractions and cell motility during development (Bresnick 1999). Third, at scaffold 7:1134324 a SNP in a *rho gtpase-activating protein* gene (hxAUGup1s7g107t1) is associated with variation in 59 DEGs associated with post-attack max swimming speed (Figure 2.9C). The activity of such G protein-activating and –deactivating gene products are essential to extracellular signaling and neuronal morphogenesis (Moon & Zheng 2003). Furthermore, Rho GTPase is one of three proteins thought to be involved in direct phosphorylation of light chain myosin II (Bresnick 1999), further supporting the association of both examples with variation in *Daphnia* swimming behavior variation. The biological functions of the genes in which the markers occur support the association with trait variation for these examples, but it is important to recognize that these, like the systems genetic hypothesis of Figure 8, should be considered hypotheses to be tested experimentally in future research.

Building on the GWA I identified numerous marker-gene-trait pathways that together suggest a systems genetic hypothesis for *Daphnia* behavioral variation. An examination of the hypothesis, which I have represented using markers occurring in genes annotated as transcriptional regulators, highlights two interesting candidates. The first, at scaffold 17:1051920, is embedded in an ecdysone-activated protein: ecdysone is a

steroid hormone central to controlling molting and development, and the sleep cycle is altered in *Drosophila* ecdysone receptor mutants (Ishimoto & Kitamoto 2010). A second gene, for which there are two markers (scaffold 183:151997 and 152040) is the MAPK gene, ERK-a. The MAPK/ERK pathway is well-established as a signaling cascade that propagates extracellular signals and stimuli to drive transcriptional responses (Seger & Krebs 1995). The connection of these (and the other) allelic variants in *Daphnia* to behaviors regulated by diel period (phototaxis), size (swimming speed), and the stimuli of vertebrate and invertebrate predators suggests fruitful avenues for future research.

This is not the first study to use exploratory genome-wide marker and transcriptomic approaches for discovering genes and loci associated with behaviors. A similar study (Ayroles et al. 2009) examined the molecular basis of locomotor reactivity and copulation latency in *Drosophila*, the former trait which appears analogous to *Daphnia* post-attack swimming velocity. Interestingly, only three genes whose expression was associated with post-attack swimming speed possess orthologs to genes associated with locomotor reactivity in *Drosophila*: hxAUG25s115g186t1 (CG5815), which is inferred in fly to be involved in cell communication and metabolic processes (FlyBase); hxAUG26up1s6g99t1 (CG12163), which is known in fly to be associated with development (Bronstein et al. 2010); and hxJGI_V11_304781 (CG33096), for which there is no functional annotation data. This relative paucity of overlap suggests that either different core mechanisms underlie variation in these similar traits, or, even though the assays are similar, they are measuring fundamentally different organismal traits. Under either scenario, the expansion of genomic approaches to *Daphnia*—some 300 million years diverged from *Drosophila* —provides new candidate genes for variation in response to mechanical stimulus.

The ecological implications of my results are varied. Swimming, whether at small or large scales, is energetically costly (Dawidowicz & Loose 1992; Loose & Dawidowicz 1994); the enrichment of genes with metabolism-related terms from expression-trait analyses recovers this aspect of *Daphnia* ecology. As expected given the role that the examined behaviors play in mediating interactions with predators, genes associated with response to stress and signaling are over-represented in both micro- and macro-scale trait classes. The loci and genes I have identified here are candidates for experimental manipulation—either by genetic crosses (Cristescu et al. 2006), RNAi (Kato et al. 2011), or related methods (Pfrender 2012)—that can be placed in the context of, e.g., classic mesocosm experiments (Hall et al. 2004; Pantel et al. 2011) to test the causal chain from sequence to ecological communities.

CONCLUSION

Using recently developed genomic resources for a classic ecological model organism I provide a foundation for future investigations into the molecular basis of variation in *Daphnia* behaviors. In addition to support for expected gene- and pathway-trait relationships I identified numerous relationships between functionally un-described genes and both macro- and micro-scale behavioral variation. Furthermore, I was able to develop an initial "systems genetic" hypothesis connecting genetic variants to differentially expressed genes that are strongly correlated with variation in nine of my focal traits. These results provide initial insight into the molecular biology of ecologically important behavioral variation, and with the accompanying supplemental resources, provide the research community in general (and the *Daphnia* genomics

community in particular) a plethora of intriguing results and hypotheses to be tested in the future.

METHODS

Clone collections and maintenance

The *D. pulex* species group is characterized by two nominal species, *D. pulex* and *D. pulicaria*, which inhabit shaded, fishless ponds and sunny, fish-filled lakes, respectively. In addition, *D. pulex* and *D. pulicaria* regularly hybridize to form lineages that are characteristic of sunny, fishless ponds (Hebert & Crease 1983). I consider the three ecotypes as *D. pulex* hereafter both for brevity and because phylogenetic (Adamowicz et al. 2009; Cristescu et al. 2012) and breeding data (Heier & Dudycha 2009) suggest that the parental "species" are not fully reproductively isolated. Furthermore, my analyses suggest weak population structure in the panel, which contains clones from each of the habitats (SI Text 1).

Nominal *D. pulex*, *D. pulicaria*, and hybrids between the two were collected from waterbodies around Michigan State University's Kellogg Biological Station, Michigan, USA, in April 2009. Individual clones were isolated, transported to the University of Texas at Austin (UT), propagated asexually in a 20°C environmental chamber under 16:8 light:dark conditions in ADaM medium (Klüttgen et al. 1994) while being fed Shellfish Diet (Reed Mariculture, Campbell, California) supplemented with live *Scenedesmus acutus*. Thirty-five clones survived the initial transfers, but three were lost before phenotyping and molecular work.

Phenotyping

Three replicates of each clone were assayed for variation in 18 swimming traits (Table 2), including macro-scale phototaxis and micro-scale swimming behaviors. The phototaxis assay was informed by the design of de Meester and Cousyn (1997); the goal was to quantify variation in phototactic behavior with and without chemical cues (kairomones; (Loose et al. 1993)) from fish predators. A six-column manifold was created from high-silica glass and assembled with silicon caulk, with 30-cm long, 5-cm inside diameter columns (SI Figure 2.4). At 06:00h local time each assay day, three randomly-chosen manifold tubes were filled with either clean ADaM medium or ADaM medium in which four fathead minnow (Pimephales promelas) were maintained in the lab; 2ml Shellfish Diet at 2x106 cells/ml was added to each manifold tube for foo*D*. Three randomly selected *Daphnia* clones from the working cultures were chosen and 20 individuals were assigned randomly to two tubes, one with and one without kairomones. The manifold was placed in a box with a hinged front and lid, which was closed to provide dark conditions. At 09:00h and 12:00h, the box lid was opened and a 5000°K fluorescent bar light was used to illuminate the manifold from above for 15 minutes; at the 15-minute mark, the front was opened and the number of *Daphnia* in the lower 4cm (L), middle 12cm (M), and upper 14cm (U) of each column was recorde*D*. This data was used with De Meester's (1996a) equation,

$$PI = (L + U) / (L + M + U)$$

to calculate the phototaxis index (PI) of each clone in each condition. The 09:00h data was used to check that PI was monotonic over time and only the 12:00h (final) PI was used in further analyses.

Micro-scale swimming behaviors were assayed in a 10 x 10 x 1cm tank made of plate glass and steel spacers and filled with ADaM and Shellfish Diet (SI Figure 2.4B). A 20-mm bar was placed on the front-right edge of the tank for scaling purposes during analysis. All assays occurred between 11:00 and 13:00h local time to match the collection times of the RNA-seq samples (see below) and remove circadian bias. An assay started by selecting a random *Daphnia* clone from the panel and gently pipetting an individual into the observation tank, then based on observational trials at different durations of pre- and post-movement behavior, allowing the individual 10 minutes to acclimate. At the 20-minute mark, video recording with a SONY HD camcorder commenced and swimming without stimulus was recorded for five minutes. At the 15- and 17.5-minute marks the observer (JM) simulated an "attack" by prodding or closely approaching the individual with a probe three times in quick succession. Video recording continued to the 20-minute mark, at which point the test individual was discarded to avoid double-sampling.

The video of each replicate was analyzed using the following procedure. First, Kinovea (www.kinovea.org) video analysis software was used to track each individual in three to five pre- and post-attack sessions with human guidance of the tracking start- and stop-points. Times when the individual was within 1.5mm of the sides or bottom were not used to avoid edge artifacts. The tracks were exported to XML files, which were parsed using custom Python scripts to extract mean, maximum, and median velocity; the coefficient of variation in velocity; the area occupied per unit time; and the sinuosity of the path, for the pre- and post-attack portions as well as the difference between the two (i.e., measuring the degree of reaction). Preliminary tests suggested discordance between

mean and median velocity, but the full data was highly correlated (r ~ 0.9), so median velocity was dropped from further analysis.

Broad-sense heritability (H^2; Lynch & Walsh 1998) and the constituent variance components (V_G and V_P), as well as inter-trait genetic and phenotypic correlations, were calculated in R 2.14 for all traits. Genetic variances and correlations were calculated from the clone means (i.e., breeding values), whereas phenotypic variances and correlations were calculated from the full data.

Nucleic acid collections

We collected both genomic DNA (gDNA) and RNA samples from each clone. gDNA was collected from pools of three individuals of each clone to obtain sufficient starting material (> 1.2μg) for 2b-RAD genotyping (Wang et al. 2012). Qiagen DNeasy kits were used for extractions per manufacturer's instructions (including RNase treatment) except on two points. First, vortexing was not used at any step in order to maximize the integrity of the gDNA; second, the final elution used only 40μl of nuclease-free water (NFW), rather than 100μl, to increase final gDNA concentrations.

To minimize expression biases due to environmental causes, RNA was collected as follows. First, a common garden was established with three replicates of three individuals of each clone grown in 150ml ADaM with daily feeding with Shellfish Diet (1ml) and twice-weekly supplements of live *Scenedesmus acutus*. Cups were cleaned, and media and food replaced, once every three days. Each cup was checked each day for reproduction, and as offspring were produced, the adults and all but five offspring were removed for three generations to remove maternal or grand-maternal effects of different clone densities in the working cultures. On two subsequent days when the final

generation of clones were 7-8d old, three individuals from each replicate were collected between 11:00-13:00h by quickly pipetting an individual into a tissue tube, removing the excess water, and immediately submerging the tube in liquid nitrogen. A small number of individuals had begun to produce ephippia, and those collection tubes were marked as such and not used for RNA extraction. Samples were quickly transferred from liquid nitrogen to -80oC freezer at the end of each collection day.

For RNA extraction, three samples of each clone were transferred from the -80°C freezer to a rack submerged in liquid nitrogen to prevent any thawing and degradation. The first two buffers of the Qiagen RNeasy extraction kits were added directly to the tubes in the liquid nitrogen bath, and tissue grinding with plastic pestles commenced immediately upon removal of the tube from the bath. Subsequent steps followed the manufacturer's instructions, including DNase addition, except that a 40μl final elution, rather than 100μl, was used to increase RNA concentrations. Note that each extraction was from a single *Daphnia*, which provided sufficient RNA for library preparation.

Genotyping

Once the samples had been sequenced, I trimmed and filtered the reads using custom Python scripts, then converted to fastq format using the SHRiMP fa2fq utility program. Next, reads were aligned to the *D. pulex* genome reference using SHRiMP 2.2.3 (David et al. 2011) with the settings –N 24 –o 10 –Q –qv-offset 33. The resultant SAM files were further processed in SAMtools (filtering out reads with MAPQ < 20, sorting; (Li et al. 2009a)) and Picard (read group additions; http://picar*D*.sourceforge.net) before genotyping with the GATK (McKenna et al. 2010). After indel realignment, an initial round of genotyping with GATK's UnifiedGenotyper (flags –ploidy 2, -glm

BOTH, -out_mode EMIT_ALL_CONFIDENT_SITES) was conducte*D*. From this I extracted variant sites from the 90th percentile of genotype qualities (GQ) and QUAL scores. Given the high-quality variants I performed base quality recalibration with the GATK's BaseRecalibrator and then performed a second round of genotyping with the UnifiedGenotyper (same setting as above). After genotyping, high-quality variants (GQ > 30, $p < 0.001$) were extracted and stored in a tab-delimited (locus x clone) file. Subsequent analyses used a reduced set of markers where at least 78% of all clones were typed at a locus. All mapping and genotyping were performed on the Stampede and Lonestar supercomputing clusters of the Texas Advanced Computing Center.

Two hazards to genotyping from expression sequences are allele-specific expression (if a locus is heterozygous but only one allele is expressed then a homozygous genotype may be incorrectly called; (Knight 2004)), and RNA editing (post-transcriptional modification of sequence; Li et al. 2009). To test if either of these occurred, I genotyped the samples using each data source independently, then compared genotypes at overlap sites and extracted any incongruities. On average, each clone was typed given RAD and RNA-seq data at 7122 loci, and possessed 10.3 differences on average (0.14%). These differences included an equal number of heterozygotes called in RAD or RNA but not vice-versa, suggesting that allele-specific expression or RNA editing were possible but very minor issues. I judged that, given the lack of population genetic data for *Daphnia* and the very low error rate, using the combined RAD and RNA data was a greater benefit than the cost of very few false-positive genotype calls.

Expression quantification

For expression quantification, I aligned the RNA-seq data to a modified version of the *Daphnia* reference genome (http://genome.jgi-psf.org/Dappu1/download/*Daphnia_pulex*.fasta.gz) that included only the 300bp immediately 5' of the annotated end of each gene reading frame and 600bp downstream (from the *Daphnia* genes 2010 GFF; http://server7.wfleabase.org/prerelease4/gene-predictions/*Daphnia*_genes2010_beta3.gff.gz), the region from which all reads should originate while permitting variants with slightly premature or extended transcription termination. The goal with this approach was to minimize mis-mapped reads (e.g., outside open reading frames) while permitting some variation in the location of mapped reads (e.g., premature or late stops). Mapping was done with SHRiMP 2.1.1b (David et al. 2011), and used the probcalc program of SHRiMP with output in the classic SHRiMP format. During post-mapping processing, a mapping was considered unique if the normalized probability was > 0.66 (for an odds ratio of ≥ 2 compared to the next-best possible mapping).

Multi-mapped reads are a problem for RNA-seq because sequence generated from single-copy genes will (should) map uniquely, while reads from genes with paralogs will multi-map more often and, if discarded or otherwise ignored, result in underestimates of expression relative to single-copy genes . Methods exist for modeling multi-mapped reads from total RNA sequencing methods (e.g., BM-map; (Ji et al. 2011)), but not for 3'-tag methods such as that used here. I experimented with using the genotype data to improve differentiating mappings between paralogs (with some, but insufficient, success); ultimately I followed one of the analyses of Taub et al. (2010) and apportioned multi-mapped reads to their targets according to a combination of the proportion of uniquely mapped reads to each mapping site (when available) and the normalized odds of

the mapping. For example, if a read mapped to two genes with normodds = 0.5, but neither gene possessed uniquely-mapped reads, then each gene read count was incremented by 0.5. If, however, one of the pair of genes possessed 9 uniquely mapped reads and the other only one, then the first gene was incremented by 0.9 and the second only 0.1. (The weightings are normalized by the sum of the products of normodds and proportion of unique reads.) To make expression comparable between samples with different coverage, values were normalized to reads per million mapped (and rounded to the nearest integer; SI Code 5). I dropped twelve samples from further analysis because they contained < 50000 mapped reads, but all clones retained at least two replicates for statistical analysis. Low expression-count genes are susceptible to inflated expression estimates from apportionment (Taub et al. 2010), but I excluded genes with an average expression < 5 reads per million mapped from further analysis, minimizing this source of bias.

To assess differential expression among clones, I fit the model

$$\text{Exp} \sim \text{clone} + \text{error}$$

in a generalized linear model (GLM; McCullagh & Nelder 1999) with quasipoisson errors and the inflation factor estimated from the pooled data (i.e., across all clones) for each gene. Significance was assessed with a likelihood ratio test against the null model, and Benjamini-Hochberg false discovery rate (FDR) control exercised at the 0.01 level.

Marker-trait and expression-trait associations

We next sought to associate polymorphic loci and differentially expressed genes (DEGs) to variation in the swimming traits. First, I reduced the marker set to include only those markers with two alleles in the population and at least three clones possessing

the minor allele (because not all clones were typed at a locus, the minor allele frequency was 9-12%). I used a linear model to test for trait ~ marker associations, where trait values were the clone means for each trait including expression (DEGs) and organismal traits. Significance was set at $p < 1e^{-7}$ for expression data and $1e^{-5}$ for organismal traits to balance association discovery with the number of tests performed. Because the large number of tests required, FDR correction rendered all relationships non-significant, and because of the novelty of the system I were more concerned with false-negatives than false-positives and used nominal p-values.

The relationship between expression need not be linear nor monotonic, and Pearson's or Spearman's correlations may miss biologically significant relationships. I tested MIC (Reshef et al. 2011) as an alternative but found that my relatively small sample size (n = 32 for clones means) was very low-powered for the method. Instead, I used distance correlation (dcor; (Székely et al. 2007)), which is not dependent on a linear or monotonic relationship between variables, to quantify the strength of association between expression variation of a DEG and variation in a trait (again, using clone means for each trait). I conducted extensive randomization experiments to estimate equations relating dcor to p¬-values. After quantifying the dcor for each gene-trait pair I exerted FDR control at a 1% rate. Genes whose expression remained significant after FDR control were considered the first set of candidates contributing to variation in each trait.

Coexpression networks, candidate refinement, and GO term and motif enrichment

After the initial candidates were identified from distance correlation, I used WGCNA (Langfelder & Horvath 2008) to (a) identify coexpression networks associated with variation in each trait and (b) remove genes whose initial association was likely

spurious. I clustered on the residuals of the trait-expression regression; genes that did not cluster with any others in the coexpression network analysis are inferred to be false-positives. Because the association of gene-trait expression was determined with dcor, which does not require linear relationships between variables, I used three regressions for each gene-trait pair (linear, log-limited, and negative exponential) and chose the residuals for the model with the highest R^2. I followed the recommended workflow produced by the authors of WGCNA (http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/), first checking for outliers, then determining the exponent required to approach a scale-free degree distribution. Modules were calculated given the exponent and adjusting the reassignThreshold and mergeCutHeight to best-match the modules apparent in the plot. Other settings included pamStage=FALSE, TOMType="unsigned", TOMDenom="mean", and minModuleSize=8. Genes assigned to the "gray" module do not cluster with any of the other genes under consideration, were considered false-positives and excluded from further analysis, leaving the final candidate set. I exported the data to Cytoscape (www.cytoscape.org) format and developed undirected graph networks for subsets of the best-connected genes by adjusting the export threshold.

After identifying the final candidate gene sets for each trait, I performed three sets of pathway/term enrichment. First, I calculated Gene Ontology (GO) enrichment using the GOstats package for R (Falcon & Gentleman 2011) for the entire set of final candidate genes. I defined two gene universes, one for all *Daphnia* genes with GO annotations and one for DEGs with GO annotations. I tested for enrichment of Molecular Function, Biological Process, and Cellular Component domains for each trait, and, because relatively few genes possess annotations, set statistical significance (pvalueCutoff) at 0.1. Next, I tested for enrichment of each coexpression module

associate with each trait to estimate the functional relevance and coherence of each module. I used REVIGO (Supek et al. 2011) to simplify and visualize GO term enrichment in the analyses.

If expression of the genes in modules is driven by one or a small number of common regulators, then I expect to find the promoter regions enriched for a concise set of binding site motifs. The promoter region (1000bp upstream and 200bp downstream of each gene's transcription start site) was extracted for each gene from the *Daphnia* reference genome using a custom script. The set of promoters for all genes within each module of each trait was tested for motif enrichment using XXmotif (Hartmann et al. 2013), using a search on the plus and minus strands, an E-value threshold of 0.1 (note that XXmotif tends to over-estimate the E-value by an order of magnitude for the average number of genes in the modules found here; Hartmann et al. [2013] Supplemental Information), and a background model of order 2. I then searched TOMTOM (meme.nbcr.net/meme/cgi-bin/tomtom.cgi, Tanaka et al. 2011) for the motifs with the strongest support to determine if known (e.g., from ChIP-seq studies) were included in the candidate motif set.

Last, I examined the final candidate genes for three classes of genes. First, I extracted genes with annotations containing the pattern [Tt]ranscription, in particular with a focus on identifying potential transcription factors as candidates driving the coexpression of genes in each module. Second, I searched for kinases, MAPKs, and signaling annotations because of their known role in transcriptional regulation. Third, I explored genes with no functional annotation using BLASTP against the UniProt manually curated protein set, and in conjunction with coexpression data, sought to

determine if some unknown genes could be annotated given a combination of data sources.

To identify putative linkages from genetic variation to behavioral variation, I identified the intersection of DEGs for which expression variation was significantly predicted by SNPs, DEGs associated with variation in each trait (Coexpression and candidate refinement), and markers significantly predictive of trait variation. my goal in requiring transitivity across levels of organization was to reduce false-positives by using multiple independent sources of information to infer relationships from the genetic to organismal trait levels. For visualization, I filtered the set of markers to include only those within genes whose annotation contained the pattern [Tt]ranscription, then extracted the DEGs and traits for which variation was strongly predicted by the allelic variants.

FIGURES

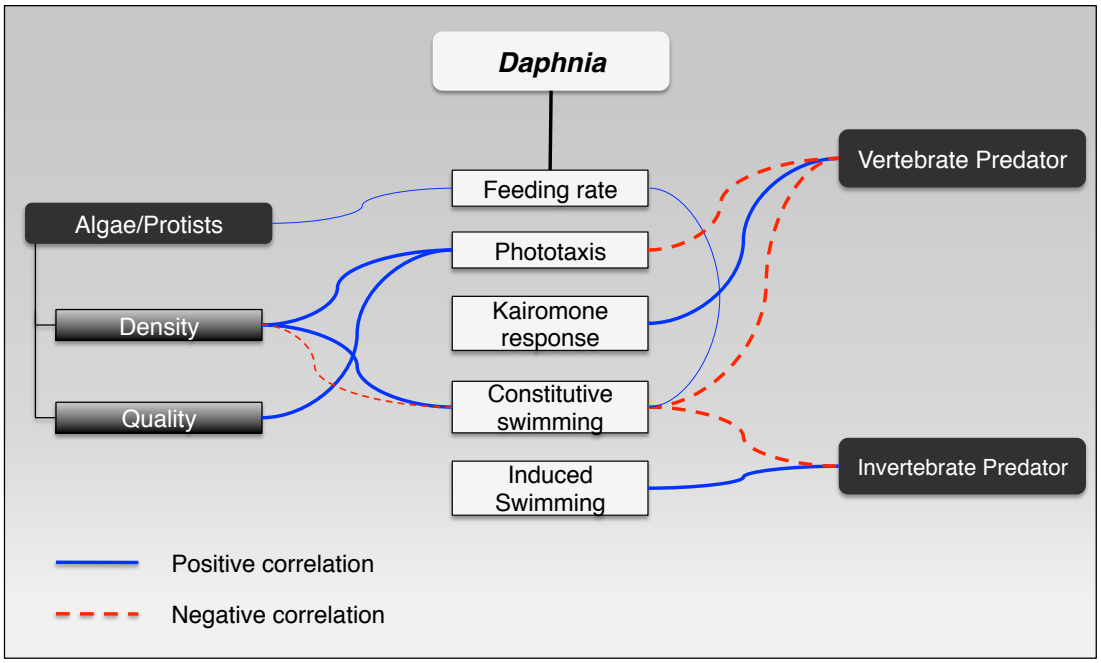


Figure 2.1. Heuristic overview of the ecological mappings of *Daphnia* behaviors. Weight of the lines connecting behaviors (white boxes, center) to aspects of the ecology (dark boxes) indicates the relative strength of the relationship. See text for further discussion.

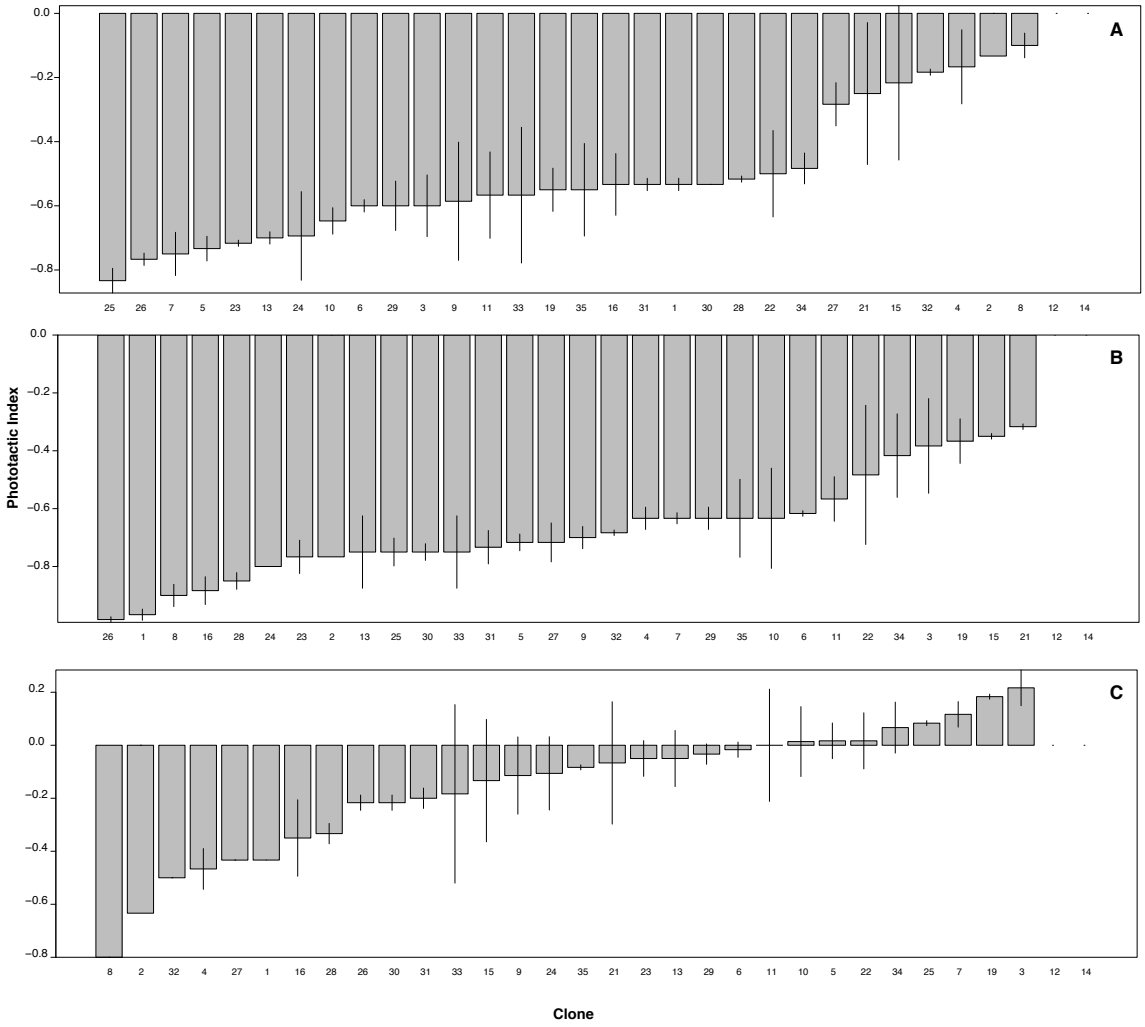


Figure 2.2. ***Daphnia*** **variation in phototaxis.** (A) The constitutive phototaxis of each clone, i.e., when swimming undisturbed in standard medium. (B) Phototaxis of clones when exposed to fish (*Pimephales promelas*) kairomones. (C) Plasticity of phototaxis, i.e., the difference between constitutive and kairomone-exposed phototaxis.

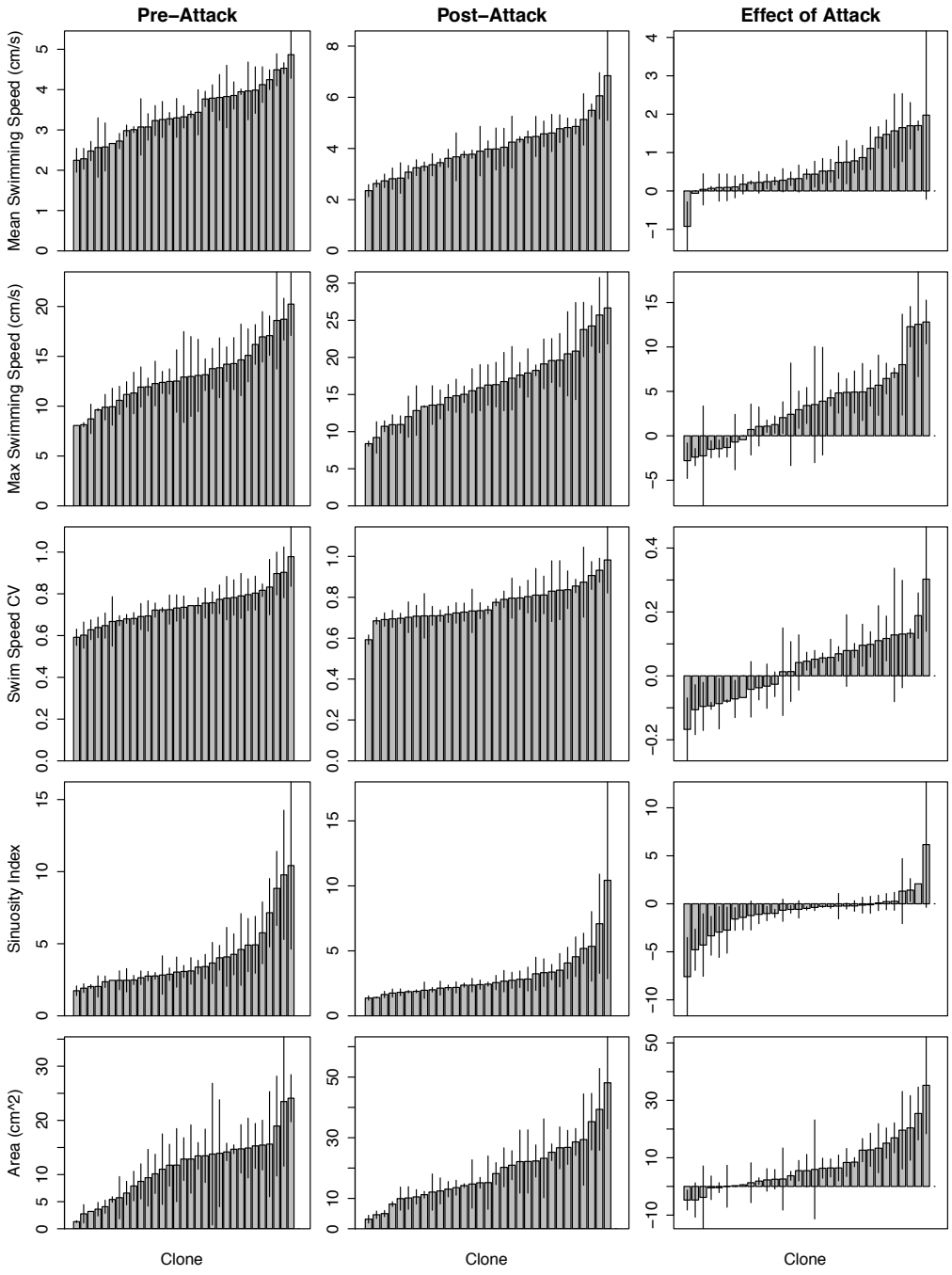

Pre−Attack
Post−Attack
Effect of Attack
Mean Swimming Speed (cm/s)
Max Swimming Speed (cm/s)
Swim Speed CV
Sinuosity Index
Area (cm^2)
Clone
Clone
Clone

Figure 2.3. ***Daphnia*** **variation in micro-scale swimming behaviors.** Each clone was measured before and after a simulated attack, and the plasticity of the behavior (Effect of Attack) is given as the difference of the pre- and post-attack values. Clone identifiers are suppressed to improve readability, but a table of clone-wise results is present in SI Table 2.1.

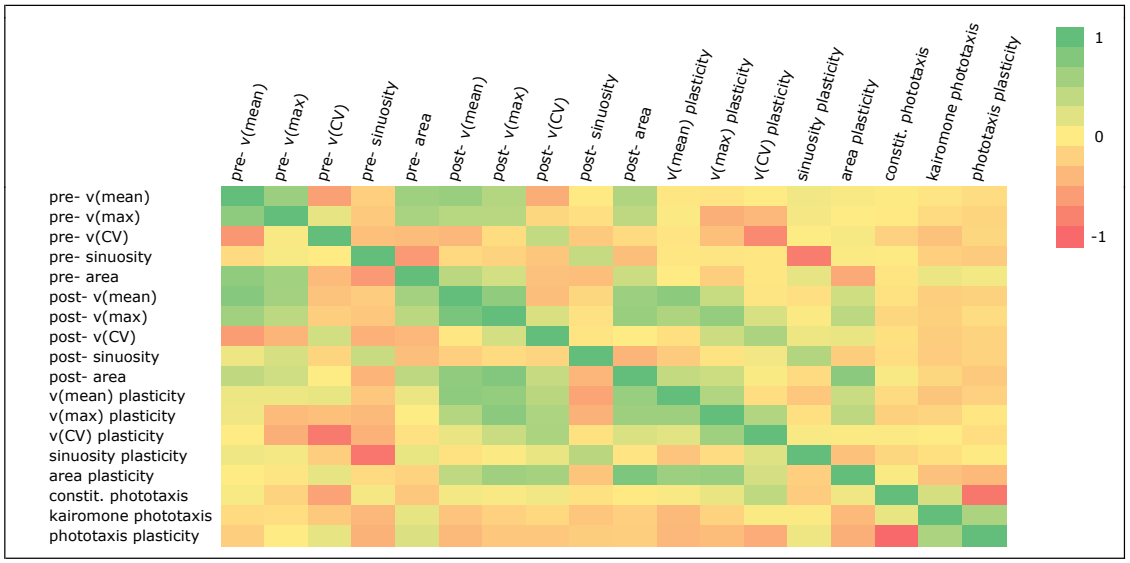


Figure 2.4. Inter-trait correlations of *Daphnia* behavior traits. Genotypic correlations are provided below the diagonal, while phenotypic correlations are provided above the diagonal. Note that phenotypic correlations tend to be weaker than genotypic correlations.

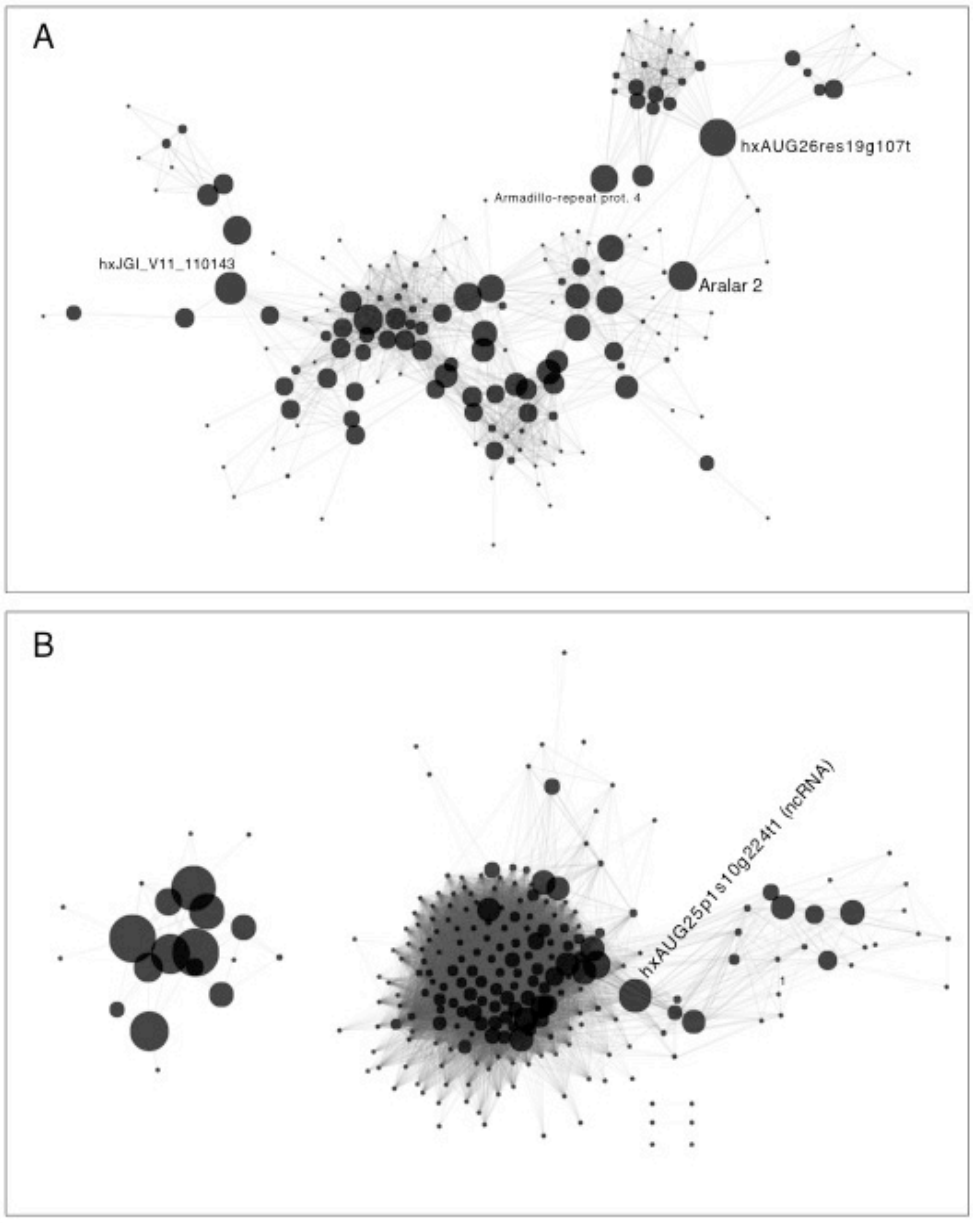


Figure 2.5. Coexpression networks associated with (A) constitutive phototaxis and (B) kairomone-exposure phototaxis. Node size is proportional to the betweenness-centrality of the gene in the network; note, however, that the value is calculated with respect to the connected subnetworks, so that the node sizes of the disconnected subnetwork on the left side of panel B is not comparable to the node sizes on the main subnetwork, right. Select high betweenness-centrality nodes are labeled and referenced in the text.

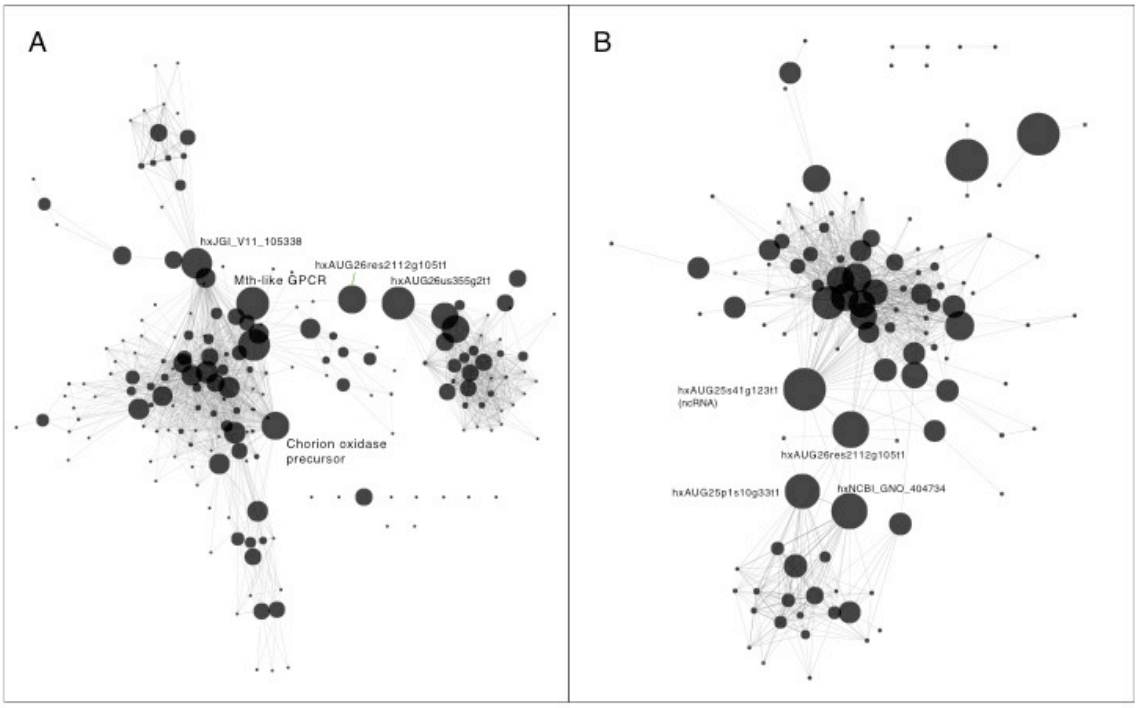


Figure 2.6. Coexpression networks associated with (A) pre-attack maximum swimming speed and (B) post-attack maximum swimming speed. Node size is proportional to the betweenness-centrality of the gene in the network; note, however, that the value is calculated with respect to the connected subnetworks, so that the node sizes of the disconnected subnetworks in the upper right-hand corner of panel B are not comparable to the node sizes in the main network. A select set of nodes with the highest betweenness-centrality are labeled and discussed in the text.

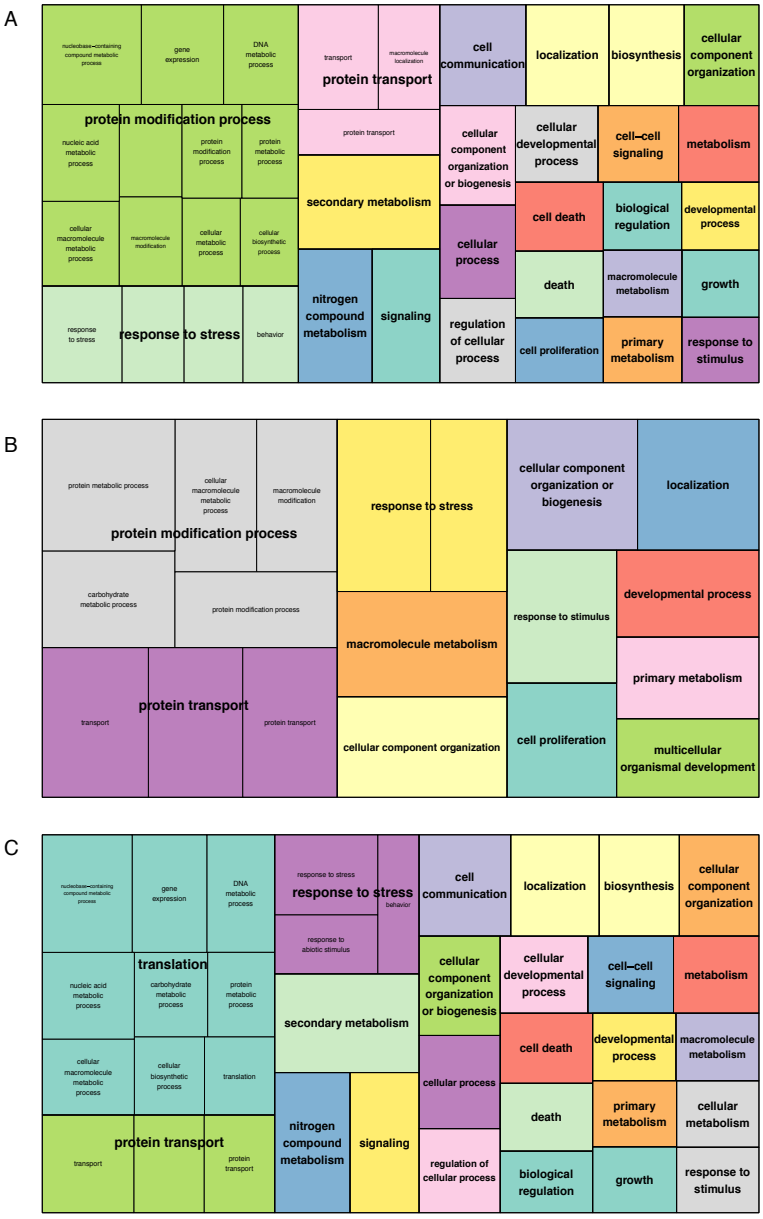


Figure 2.7. Gene Ontology Biological Process enrichment across (A) all behavioral traits, (B) phototaxis traits, and (C) micro-scale swimming traits. Each panel is based on the reduction and summary analysis produced by REVIGO (Supek et al. 2011), given the enrichment calculations from GOstats (Gentleman and Falcon 2011). Large-font labels are representative terms of the color-coded blocks over which they are found, and block size reflects the p-value of the term.

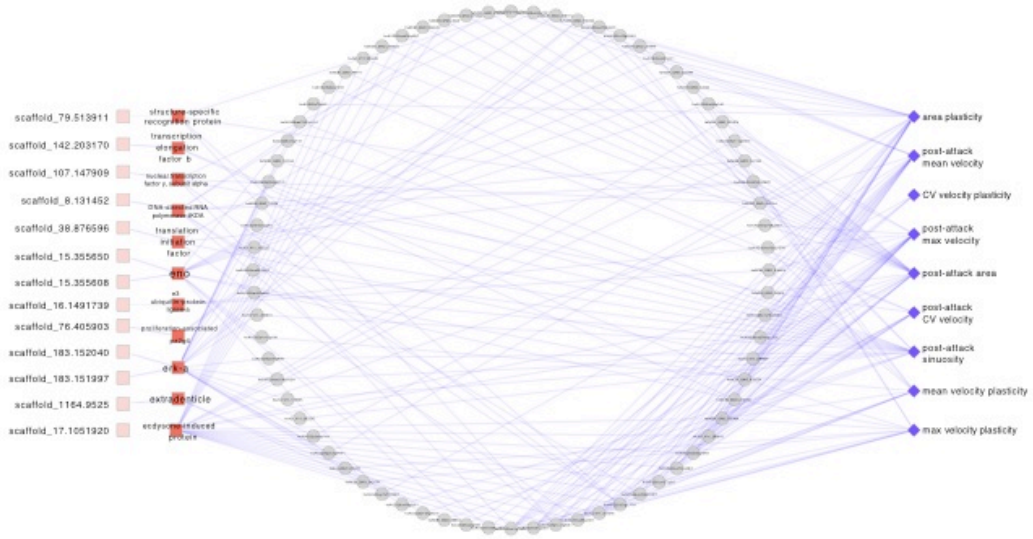


Figure 2.8. An integrated gene network hypothesis for *Daphnia* behaviors. Genetic variants and transcriptional regulators in which they are found are represented by red-shade squares; differentially expressed genes (DEGs) whose expression is predicted by allelic variants, and which are correlated with behavioral variation, are represented by gray circles; and the behavioral traits to which these map are represented by blue diamonds. The labels of the DEGs, and inter-DEG correlations, are suppressed to improve clarity; the data are in SI Table 2.7.

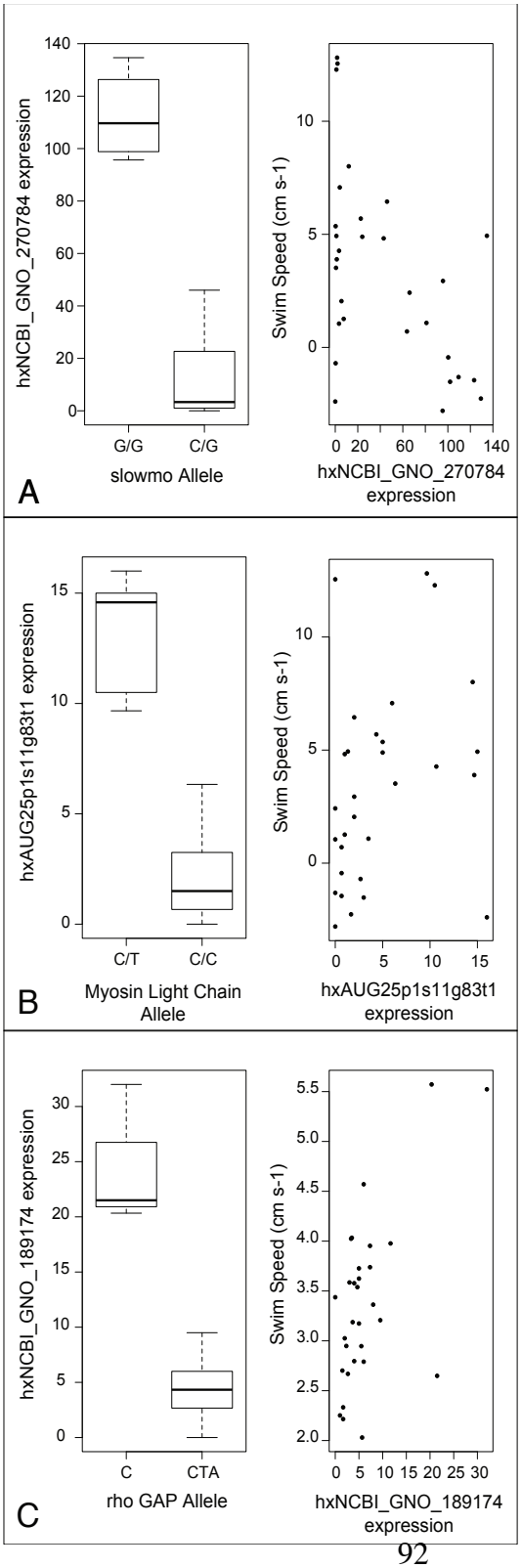

hxNCBI_GNO_270784 expression
140
120
100
80
60
40
20
0
G/G
C/G
slowmo Allele
A
Swim Speed (cm s-1)
10
5
0
0 20 60 100 140
hxNCBI_GNO_270784 expression
hxAUG25p1s11g83t1 expression
15
10
5
0
C/T
C/C
Myosin Light Chain Allele
B
Swim Speed (cm s-1)
10
5
0
0 5 10 15
hxAUG25p1s11g83t1 expression
hxNCBI_GNO_189174 expression
30
25
20
15
10
5
0
C
CTA
rho GAP Allele
C
Swim Speed (cm s-1)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0 5 10 15 20 25 30
hxNCBI_GNO_189174 expression

Figure 2.9. Box- and scatterplots relating allelic variants to gene expression, and gene expression to behavioral variation. (previous page) (A) The two alleles in *slowmo* are strongly related to expression of an un-annotated gene whose expression is negatively correlated with post-attack maximum swimming speed. Note that two expression-trait patterns are present, suggesting a biological interaction with other gene(s). (B) The *nonmuscle myosin light chain* alleles are predictive of expression of an un-annotated genes whose expression is positively correlated with post-attack maximum swimming speed. (C) *Rho GAP*, whose activity is associated with regulating myosin light chain, possesses two alleles (one of which is a two-base insertion) that are strongly associated with expression of an un-annotated gene that is positively correlated with post-attack swimming speed. As discussed in the text, these relationships are statistically supported and should be considered hypotheses to be experimentally tested.

TABLES

Trophic interaction		Group	Ecological Context	References
competition	*predation*			
X	X		increased consumption of high-P seston, decreased exposure to invertebrate predators, increased susceptibility to vertebrate predators, variation in competition with heterospecifics	Boersma et al. 1998; Cousyn et al. 2001; de Meester 1991; de Meester et al. 1996, 1997, 1999, 2008; Dawidowicz and Loose 1992; Dodson 1988; Duffy 2010; Gliwicz and Maszczyk 2007; Doksaeter and Vijerberg 2001; Hanazato and Dodson 1995; Hood andSterner 2010; Kerfoot 1980; Lampert 2006; Lampert and Taylor 1985; Leibold et al 1994; Loose and Dawidowicz 1984; Loose 1993; Mikulski and Pijanowska 2010; Pauwels et al. 2005; Reede 2003; Tessier and Leibold 1997; van Gool and Ringleberg 1998; Winder et al. 2004
X	X		decreased susceptibility to vertebrate predators, increased susceptibility to invertebrate predators, interaction with oxygen availability	
	X		balance between food intake and predator exposure	
X	X		increased food acquisition, probability of vertebrate predator detection, invertebrate predator detection, disease susceptibility; decreased phytoplankton/protist/bacterial density, interaction with body size	Boeing et al. 2006; Brewer and Coughlin 1996; Brewer et al. 1999; Dodson and Ramcharan 1991; Dodson and Hanazato 1995; Dodson et al. 1997; Drenner et al. 1978; Gliwicz and Maszczyk 2007; Hall et al. 2007; Kerfoot 1980; O'Keefe et al. 1998; Pijanowska et al. 1993; Porter et al. 1982; Sarnelle 2005; Seuront et al. 2004a,b; Ware 1973; Weber and Noordwijk 2002
	X			
	X			
	X			
X	X			
X	X		decreased susceptibility to invertebrate predators, increased food uptake, improved response to turbulence,	
	X			
	X			
	X			
X	X			
X	X		plastic response to invertebrate predators, subsequent but temporary effects on foraging rate,	
	X			
	X			
	X			
X	X			

Table 2.1. Relationships between the 18 *Daphnia* behavioral traits and ecological context, with supporting references. An 'X' indicated a mapping to the expected trophic interaction. Traits are grouped according to macro-scale behaviors and pre-attack, post-attack, and deviation (post – pre, labeled "plasticity") behaviors.

Category	Class	Phototaxis	Kairomone	Pre-attack swimming	Post-attack swimming
Metabolism	Carbohydrate metabolism	■	■	■	■
	Lipid metabolism	■			
	Nucleotide/amino acid metabolism	■			
	Glycan biosynthesis				
	Secondary metabolite biosynthesis				
Genetic Information Processing	Transcription	■	■	■	■
	Translation: Ribosome	■	■	■	■
	Folding, sorting, degradation				
Environmental Information Processing	Membrane transport		■		
	Signal transduction: mTOR				
	Signal transduction: MAPK	■	■	■	■
	Signaling molecules: G protein-coupled receptors	■	■	■	■
	Signaling molecules: Ion channels	■	■	■	■
	Signaling molecules: Cytokine receptors		■	■	■
Cellular Processes	Transport and catabolism				
	Cell growth and death: cell cycle				
	Cell growth and death: apoptosis				
Organismal Systems	Immune system: Toll-like receptor signaling		■	■	■
	Endocrine system: Insulin signaling pathway		■	■	■
	Digestive system: carbohydrate, protein, fat digestion	■			
	Sensory system: phototransduction	■	■		
	Sensory system: taste transduction		■		

Table 2.2. Expected relationships between groups of *Daphnia* behavioral traits and groups of biological terms derived from Gene Ontology and KEGG terms. We anticipate recovering genes with the term enrichments for the trait combinations with blue blocks.

Trait	H^2	CV_g	V_g	V_p	mean	CV_r
phototaxis	0.771	-0.088	0.046	0.060	-0.523	22.399
fish kairomone phototaxis	0.802	-0.056	0.037	0.046	-0.669	14.321
fish kairomone plasticity	0.794	-0.444	0.065	0.082	-0.146	88.766
pre-attack mean swim speed	0.710	0.171	0.588	0.828	3.439	14.238
pre-attack max swim speed	0.588	0.942	12.465	21.201	13.234	22.334
pre-attack swim speed CV	0.573	0.015	0.011	0.019	0.741	12.203
pre-attack swim sinuosity	0.601	1.778	7.073	11.761	3.978	54.420
pre-attack swim area occupied	0.483	3.950	45.966	95.246	11.637	60.326
post mean swim speed	0.721	0.324	1.310	1.816	4.040	17.607
post max swim speed	0.687	1.719	28.380	41.336	16.514	21.797
post swim speed CV	0.512	0.014	0.011	0.021	0.771	13.004
post swim sinuosity	0.537	1.656	5.100	9.501	3.080	68.101
post swim area occupied	0.680	7.529	141.319	207.952	18.771	43.487
mean swim speed plasticity	0.570	0.981	0.619	1.087	0.631	108.380
max swim speed plasticity	0.634	7.130	24.403	38.514	3.423	109.755
swim speed CV plasticity	0.575	0.464	0.015	0.025	0.031	330.874
swim sinuosity plasticity	0.567	-8.396	7.483	13.205	-0.891	268.354
swim area occupied plasticity	0.593	15.730	116.725	196.745	7.420	120.552

Table 2.3. Quantitative genetics of the 18 behavioral traits measured in the *Daphnia* panel.

Trait	Gene summaries				Modularity summary		
	# genes	# annotated	# unannotated	Pct. annotated	# modules	Min. module size	Max. module size
phototaxis	198	72	126	36.36	9	8	41
fish kairomone phototaxis	239	108	128	45.76	4	15	141
fish kairomone plasticity	236	102	137	42.68	8	10	60
pre-attack swim area occupied	164	78	86	47.56	3	28	84
pre-attack swim speed CV	228	86	142	37.72	8	8	64
pre-attack max swim speed	156	44	112	28.21	8	10	36
pre-attack mean swim speed	149	62	87	41.61	8	9	54
pre-attack swim sinuosity	175	57	118	32.57	2	86	89
post swim area occupied	404	155	249	38.37	7	8	235
post swim speed CV	166	70	96	42.17	6	9	81
post max swim speed	263	99	164	37.64	12	8	58
post mean swim speed	289	109	183	37.33	9	12	88
post swim sinuosity	223	71	152	31.84	1	223	223
swim area occupied plasticity	314	133	181	42.36	3	23	216
swim speed CV plasticity	172	69	103	40.12	9	13	29
max swim speed plasticity	185	82	103	44.32	9	8	49
mean swim speed plasticity	182	70	112	38.46	7	8	81
swim sinuosity plasticity	71	23	48	32.39	2	8	63

Table 2.4. Summary overview of *Daphia* behavioral trait-gene relationships.

LITERATURE CITED

Acharya, K., Kyle, M. & Elser, J.J. (2004). Effects of stoichiometric dietary mixing on *Daphnia* growth and reproduction. *Oecologia*, 138, 333–340.

Adamowicz, S.J., Petrusek, A., Colbourne, J.K., Hebert, P.D.N. & Witt, J.D.S. (2009). The scale of divergence: A phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. *Molecular Phylogenetics and Evolution*, 50, 423–436.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., *et al.* (2009). Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet*, 41, 299–307.

Barton, N.H., Briggs, D.E.G., Eisen, J.A., Goldstein, D.B. & Patel, N.H. (2007). *Evolution*. Cold Springs Harbor Laboratory Press, New York.

Bell, A.M. (2009). Approaching the Genomics of Risk-Taking Behavior. In: *Advances in Genetics*. Elsevier, pp. 83–104.

Boeing, W.J., Ramcharan, C.W. & Riessen, H.P. (2006). Multiple predator defence strategies in *Daphnia pulex* and their relation to native habitat. *Journal of Plankton Research*, 28, 571 –584.

Boersma, M., Spaak, P. & De Meester, L. (1998). Predator-mediated plasticity in morphology, life history, and behavior of *Daphnia*: the uncoupling of responses. *The American Naturalist*, 152, 237–248.

Bresnick, A.R. (1999). Molecular mechanisms of nonmuscle myosin-II regulation. *Current Opinion in Cell Biology*, 11, 26–33.

Brewer, M.C., Dawidowicz, P. & Dodson, S.I. (1999). Interactive effects of fish kairomone and light on *Daphnia* escape behavior. *Journal of plankton research*, 21, 1317–1335.

Bronstein, R., Levkovitz, L., Yosef, N., Yanku, M., Ruppin, E., Sharan, R., *et al.* (2010). Transcriptional Regulation by CHIP/LDB Complexes. *PLoS Genet*, 6, e1001063.

Carhan, A., Reeve, S., Dee, C.T., Baines, R.A. & Moffat, K.G. (2004). Mutation in slowmo causes defects in *Drosophila* larval locomotor behaviour. *Invert. Neurosci.*, 5, 65–75.

Carroll, S.B. (2009). Endless forms: the evolution of gene regulation and morphological diversity. *Philosophy of Biology: An Anthology*, 32, 193.

Christie, A.E. & McCoole, M.D. (2012). From genes to behavior: investigations of neurochemical signaling come of age for the model crustacean *Daphnia pulex*. *J Exp Biol*, 215, 2535–2544.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., *et al.* (2011). The Ecoresponsive Genome of *Daphnia pulex*. *Science*, 331, 555 – 561.

Cooper, S.D. & Smith, D.W. (1982). Competition, predation and the relative abundances of two species of *Daphnia*. *J. Plankton Res.*, 4, 859–879.

Cousyn, C., De Meester, L., Colbourne, J.K., Brendonck, L., Verschuren, D. & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, 98, 6256.

Cristescu, M.E., Constantin, A., Bock, D.G., Cáceres, C.E. & Crease, T.J. (2012). Speciation with gene flow and the genetics of habitat transitions. *Mol. Ecol.*, 21, 1411–1422.

Cristescu, M.E.A., Colbourne, J.K., Radivojac, J. & Lynch, M. (2006). A microsatellite-based genetic linkage map of the waterflea, *Daphnia pulex*: On the prospect of crustacean genomics. *Genomics*, 88, 415–430.

David, M., Dzamba, M., Lister, D., Ilie, L. & Brudno, M. (2011). SHRiMP2: Sensitive yet Practical Short Read Mapping. *Bioinformatics*, 27, 1011 –1012.

Dawidowicz, P. & Loose, C.J. (1992). Cost of Swimming by *Daphnia* During Diel Vertical Migration. *Limnology and Oceanography*, 37, 665–669.

DeMott, W.R. & Kerfoot, W.C. (1982). Competition among cladocerans: nature of the interaction between Bosmina and *Daphnia*. *Ecology*, 63, 1949–1966.

Dodson, S. (1988). The Ecological Role of Chemical Stimuli for the Zooplankton: Predator-Avoidance Behvior in *Daphnia*. *Limnology and Oceanography*, 33, 1431–1439.

Dodson, S. & Ramcharan, C. (1991). Size-specific swimming behavior of *Daphnia pulex*. *Journal of Plankton Research*, 13, 1367 –1379.

Dodson, S.I. (1974). Zooplankton Competition and Predation: An Experimental Test of the Size-Efficiency Hypothesis. *Ecology*, 55, 605–613.

Dodson, S.I. & Hanazato, T. (1995). Commentary on effects of anthropogenic and natural organic chemicals on development, swimming behavior, and reproduction of *Daphnia*, a key member of aquatic ecosystems. *Environmental Health Perspectives*, 103, 7.

Dodson, S.I., Newman, A.L., Will-Wolf, S., Alexander, M.L., Woodford, M.P. & Van Egeren, S. (2009). The relationship between zooplankton community structure and lake characteristics in temperate lakes (Northern Wisconsin, USA). *J. Plankton Res.*, 31, 93–100.

Doksæter, A. & Vijverberg, J. (2001). The effects of food and temperature regimes on life-history responses to fish kairomones in *Daphnia* hyalina$\times$ galeata*. *Hydrobiologia*, 442, 207–214.

Drenner, R.W., Strickler, J.R. & O'Brien, W.J. (1978). Capture Probability: The Role of Zooplankter Escape in the Selective Feeding of Planktivorous Fish. *J. Fish. Res. Bd. Can.*, 35, 1370–1373.

Duffy, M.A. (2010). Ecological consequences of intraspecific variation in lake *Daphnia*. *Freshwater Biology*, 55, 995–1004.

Enserink, E., Maasdiepeveen, J. & Vanleeuwen, C. (1991). Combined effects of metals - an ecotoxicological evaluation. *Water Research*, 25, 679–687.

Falcon, S. & Gentleman, R. (2011). How To Use GOstats Testing Gene Lists for GO Term Association.

Forbes, S.A. (1880). The Food of the Darters. *The American Naturalist*, 14, 697–703.

Gliwicz, Z. & Umana, G. (1994). Cladoceran body-size and vulnerability to copepod predation. *Limnol. Oceanogr.*, 39, 419–424.

Gliwicz, Z.M. & Maszczyk, P. (2006). *Daphnia* growth is hindered by chemical information on predation risk at high but not at low food levels. *Oecologia*, 150, 706–715.

Van Gool, E. & Ringelberg, J. (1998). Light-induced migration behaviour of< i> *Daphnia*</i> modified by food and predator kairomones. *Animal behaviour*, 56, 741–747.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. (2007). Quantifying similarity between motifs. *Genome biology*, 8, R24.

Hall, S., Leibold, M., Lytle, D. & Smith, V. (2004). Stoichiometry and planktonic grazer composition over gradients of light, nutrients, and predation risk. *ECOLOGY*, 85, 2291–2301.

Hall, S.R., Sivars-Becker, L., Becker, C., Duffy, M.A., Tessier, A.J. & Cáceres, C.E. (2007). Eating yourself sick: transmission of disease as a function of foraging ecology. *Ecology Letters*, 10, 207–218.

Hamilton, M. (2009). *Population Genetics*. 1st edn. Wiley-Blackwell.

Hanazato, T. & Dodson, S.I. (1995). Synergistic effects of low oxygen concentration, predator kairomone, and a pesticide on the cladoceran *Daphnia pulex*. *Limnology and Oceanography*, 700–709.

Hartmann, H., Guthöhrlein, E.W., Siebert, M., Luehr, S. & Söding, J. (2013). P-value-based regulatory motif discovery using positional weight matrices. *Genome Res.*, 23, 181–194.

Hebert, P.D.N. & Crease, T. (1983). Clonal diversity in populations of *Daphnia pulex* reproducing by obligate parthenogenesis. *Heredity*, 51, 353–369.

Hebert, P.D.N., Ward, R.D. & Weider, L.J. (1988). Clonal-Diversity Patterns and Breeding-System Variation in *Daphnia pulex*, asexual-Sexual Complex. *Evolution*, 42, 147–159.

Heier, C. & Dudycha, J.L. (2009). Ecological speciation in a cyclic parthenogen: Sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia* pulicaria. *Limnology and Oceanography*, 54, 492–502.

Hoekstra, H.E. & Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, 61, 995–1016.

Hood, J.M. & Sterner, R.W. (2010). Diet Mixing: Do Animals Integrate Growth or Resources across Temporal Heterogeneity? *The American Naturalist*, 176, 651–663.

Hülsmann, S., Rinke, K. & Mooij, W.M. (2011). Size-selective predation and predator-induced life-history shifts alter the outcome of competition between planktonic grazers. *Functional Ecology*, 25, 199–208.

Inoue, K. & Lupski, J.R. (2003). Genetics and genomics of behavioral and psychiatric disorders. *Current opinion in genetics & development*, 13, 303–309.

Ishimoto, H. & Kitamoto, T. (2010). The Steroid Molting Hormone Ecdysone Regulates Sleep in Adult *Drosophila melanogaster*. *Genetics*, 185, 269–281.

Jeyasingh, P.D., Ragavendran, A., Paland, S., Lopez, J.A., Sterner, R.W. & Colbourne, J.K. (2011). How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*. *Mol. Ecol.*, 20, 2341–2352.

Ji, Y., Xu, Y., Zhang, Q., Tsui, K.-W., Yuan, Y., Norris Jr., C., *et al.* (2011). BM-Map: Bayesian Mapping of Multireads for Next-Generation Sequencing Data. *Biometrics*, 67, 1215–1224.

Jordan, K., Carbone, M., Yamamoto, A., Morgan, T. & Mackay, T. (2007). Quantitative genomics of locomotor behavior in *Drosophila melanogaster*. *Genome Biology*, 8, R172.

Kato, Y., Shiga, Y., Kobayashi, K., Tokishita, S., Yamagata, H., Iguchi, T., *et al.* (2011). Development of an RNA interference method in the cladoceran crustacean *Daphnia* magna. *Dev Genes Evol*.

Kerfoot, W.C., Kellogg Jr, D.L. & Strickler, J.R. (1980). Visual observations of live zooplankters: evasion, escape, and chemical defenses. *Evolution and ecology of zooplankton communities. University Press of New England*, 10–27.

Klüttgen, B., Dülmer, U., Engels, M. & Ratte, H.. (1994). ADaM, an artificial freshwater for the culture of zooplankton. *Water Research*, 28, 743–746.

Knight, J.C. (2004). Allele-specific gene expression uncovered. *Trends in Genetics*, 20, 113–116.

Lampert, W. (1993). Phenotypic Plasticity of the Size at First Reproduction in *Daphnia*: The Importance of Maternal Size. *Ecology*, 74, 1455–1466.

Lampert, W. (2006). *Daphnia*: model herbivore, predator and prey. *Polish Journal of Ecology*, 54, 607–620.

Lampert, W. & Taylor, B.E. (1985). Zooplankton Grazing in a Eutrophic Lake: Implications of Diel Vertical Migration. *Ecology*, 66, 68–82.

Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.

Leibold, M., Tessier, A. & West, C. (1994). Genetic, acclimatization, and ontogenic effects on habitat selection behavior in *Daphnia* pulicaria. *Evolution*, 48, 1324–1332.

Leibold, M. & Tessier, A.J. (1991). Contrasting patterns of body size for *Daphnia* species that segregate by habitat. *Oecologia*, 86, 342–348.

Leibold, M.A. (1991). Trophic interactions and habitat segregation between competing *Daphnia* species. *Oecologia*, 86, 510–520.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009a). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., LeProust, E., *et al.* (2009b). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science*, 324, 1210–1213.

Loose, C.J. & Dawidowicz, P. (1994). Trade-offs in diel vertical migration by zooplankton: the costs of predator avoidance. *Ecology*, 2255–2263.

Loose, C.J., Von Elert, E. & Dawidowicz, P. (1993). Chemically-induced diel vertical migration in *Daphnia*: a new bioassay for kairomones exuded by fish. *Archiv fur Hydrobiologie. Stuttgart*, 126, 329–337.

Lynch, M. & Walsh, B. (1998). Genetics and analysis of quantitative traits.

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., *et al.* (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, 482, 173–178.

Main, T., Dobberfuhl, D. & Elser, J. (1997). N : P stoichiometry and ontogeny of crustacean zooplankton: A test of the growth rate hypothesis. *Limnology and Oceanography*, 42, 1474–1478.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Exploring the molecular basis of variation in size and fecundity in an emerging model organism, *Daphnia pulex*.

McCoole, M.D., Baer, K.N. & Christie, A.E. (2011). Histaminergic signaling in the central nervous system of *Daphnia* and a role for it in the control of phototactic behavior. *The Journal of Experimental Biology*, 214, 1773 –1782.

McCullagh, P. & Nelder, J.A. (1999). *Generalized linear models*. Chapman & Hall, CRC.

McGuffin, P., Riley, B. & Plomin, R. (2001). Toward behavioral genomics. *Science*, 291, 1232–1249.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20, 1297–1303.

De Meester, L. (1991). An analysis of the phototactic behaviour of *Daphnia* magna clones and their sexual descendants. *Hydrobiologia*, 225, 217–227.

De Meester, L. (1993). Genotype, fish-mediated chemical, and phototactic behavior in *Daphnia* magna. *Ecology*, 74, 1467–1474.

De Meester, L. (1996a). Evolutionary potential and local genetic differentiation in a phenotypically plastic trait of a cyclical parthenogen, *Daphnia* magna. *Evolution*, 1293–1298.

De Meester, L. (1996b). Evolutionary potential and local genetic differentiation in a phenotypically plastic trait of a cyclical parthenogen, *Daphnia* magna. *Evolution*, 1293–1298.

De Meester, L. & Cousyn, C. (1997). The change in phototactic behaviour of a *Daphnia* magna clone in the presence of fish kairomones: the effect of exposure time. *Hydrobiologia*, 360, 169–175.

De Meester, L. & Weider, L.J. (1999). Depth selection behavior, fish kairomones, and the life histories of *Daphnia* hyalina$\times$ galeata hybrid clones. *Limnology and oceanography*, 1248–1258.

Meyer, E., Aglyamova, G.V. & Matz, M.V. (2011). Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology*, 20, 3599–3616.

Mikulski, A. & Pijanowska, J. (2010). When and how can *Daphnia* prepare their offspring for the threat of predation? *Hydrobiologia*, 643, 21–26.

Miner, B.E., Meester, L.D., Pfrender, M.E., Lampert, W. & Hairston, N.G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proc. R. Soc. B*, 279, 1873–1882.

Moon, S.Y. & Zheng, Y. (2003). Rho GTPase-activating proteins in cell regulation. *Trends in Cell Biology*, 13, 13–22.

Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., *et al.* (2002). The extent of linkage disequilibrium in Arabidopsis thaliana. *Nature Genetics*, 30, 190–193.

O'Keefe, T.C., Brewer, M.C. & Dodson, S.I. (1998). Swimming behavior of *Daphnia*: its role in determining predation risk. *Journal of Plankton Research*, 20, 973–984.

Pantel, J.H., Leibold, M.A. & Juenger, T.E. (2011). Population differentiation in *Daphnia* alters community assembly in experimental ponds. *Am. Nat.*, 177, 314–322.

Pauwels, K., Stoks, R. & De Meester, L. (2005). Coping with predator stress: interclonal differences in induction of heat-shock proteins in the water flea *Daphnia* magna. *Journal of evolutionary biology*, 18, 867–872.

Penalva-Arana, D.C., Lynch, M. & Robertson, H. (2009). The chemoreceptor genes of the waterflea *Daphnia pulex*: many Grs but no Ors. *BMC Evolutionary Biology*, 9, 79.

Pereira, H.S. & Sokolowski, M.B. (1993). Mutations in the larval foraging gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 5044–5046.

Pfrender, M.E. (2012). Triangulating the genetic basis of adaptation to multifarious selection. *Molecular Ecology*, 21, 2051–2053.

Pijanowska, J., Weider, L.J. & Lampert, W. (1993). Predator-mediated genotypic shifts in a prey population: experimental evidence. *Oecologia*, 96, 40–42.

Porter, K.G., Gerritsen, J. & Orcutt, J.D. (1982). The Effect of Food Concentration on Swimming Patterns, Feeding Behavior, Ingestion, Assimilation, and Respiration by *Daphnia*. *Limnology and Oceanography*, 27, 935–949.

Reede, T. (1995). Life history shifts in response to different levels of fish kairomones in *Daphnia*. *Journal of plankton research*, 17, 1661–1667.

Renn, S.C.P., Aubin-Horth, N. & Hofmann, H.A. (2008). Fish and chips: functional genomics of social plasticity in an African cichlid fish. *J Exp Biol*, 211, 3041–3056.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., *et al.* (2011). Detecting Novel Associations in Large Data Sets. *Science*, 334, 1518 –1524.

Rhee, S.Y., Wood, V., Dolinski, K. & Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nature Reviews Genetics*, 9, 509–515.

Robinson, G.E., Fernald, R.D. & Clayton, D.F. (2008). Genes and social behavior. *science*, 322, 896–900.

Rouvalis, A., Karadima, C., Zioris, I., Sakkas, V., Albanis, T. & Iliopoulou-Georgudaki, J. (2009). Determination of pesticides and toxic potency of rainwater samples in western Greece. *Ecotoxicology And Environmental Safety*, 72, 828–833.

Sarnelle, O. (2005). *Daphnia* as keystone predators: effects on phytoplankton diversity and grazing resistance. *Journal of plankton research*, 27, 1229.

Scheiner, S.M. & Berrigan, D. (1998). The genetics of phenotypic plasticity. VIII. The cost of plasticity in *Daphnia pulex*. *Evolution*, 368–378.

Seger, R. & Krebs, E.G. (1995). The MAPK signaling cascade. *FASEB J*, 9, 726–735.

Seuront, L., Brewer, M.C., Strickler, J.R. & others. (2004). Quantifying zooplankton swimming behavior: the question of scale. *Handbook of scaling methods in aquatic ecology—measurements, analysis, simulation. CRC Press, Boca Raton*, 333–359.

Simon, J.-C., Pfrender, M.E., Tollrian, R., Tagu, D. & Colbourne, J.K. (2011). Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. *Journal of Heredity*, 102, 512 –525.

Spitze, K. (1992). Predator-mediated plasticity of prey life-history and morphology – *Chaoborus americanus* predation on *daphnia-pulex*. *American Naturalist*, 139, 229–247.

Stern, D. & Orgogozo, V. (2009). Is genetic evolution predictable? *SCIENCE*, 323, 746.

Stern, D.L. & Orgogozo, V. (2008). The Loci of Evolution: How Predictable is Genetic Evolution? *Evolution*, 62, 2155–2177.

Stone, E.A. & Ayroles, J.F. (2009). Modulated Modularity Clustering as an Exploratory Tool for Functional Genomic Inference. *PLoS Genet.*, 5, e1000479.

Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, 6, e21800.

Székely, G.J., Rizzo, M.L. & Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35, 2769–2794.

Tanaka, E., Bailey, T., Grant, C.E., Noble, W.S. & Keich, U. (2011). Improved similarity scores for comparing motifs. *Bioinformatics*, 27, 1603–1609.

Taub, M.A., Corrada Bravo, H. & Irizarry, R.A. (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome Med*, 2, 87.

Tessier, A. & Goulden, C. (1982). Estimating food limitation in cladoceran populations. *Limnology and Oceanography*, 27, 707–717.

Tessier, A., Young, A. & Leibold, M. (1992). Population dynamics and body-size selection in *Daphnia*. *Limnology and Oceanography*, 37, 1–13.

Tessier, A.J. & Leibold, M.A. (1997). Habitat use and ecological specialization within lake *Daphnia* populations. *Oecologia*, 109, 561–570.

Tilden, A.R., McCoole, M.D., Harmon, S.M., Baer, K.N. & Christie, A.E. (2011). Genomic identification of a putative circadian system in the cladoceran crustacean *Daphnia pulex*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 6, 282–309.

Wang, S., Meyer, E., McKay, J.K. & Matz, M.V. (2012). 2b-rad: a simple and flexible method for genome-wide genotyping. *Nature Methods*, 9, 808–810.

Ware, D.M. (1973). Risk of epibenthic prey to predation by rainbow trout (*Salmo gairdneri*). *Journal of the Fisheries Board of Canada*, 30, 787–797.

Weber, A. & Van Noordwijk, A. (2002). Swimming behaviour of *Daphnia* clones: differentiation through predator infochemicals. *Journal of Plankton Research*, 24, 1335 –1348.

Weismann, A. (1876). *Beiträge zur Naturgeschichte der Daphnoiden*. W. Engelmann.

Wetzel, R.G. (2001). *Limnology: lake and river ecosystems*. Academic press.

Winder, M., Spaak, P. & Mooij, W.M. (2004). Trade-offs in *Daphnia* habitat selection. *Ecology*, 85, 2027–2036.

Yarmolinsky, D.A., Zuker, C.S. & Ryba, N.J.P. (2009). Common Sense about Taste: From Mammals to Insects. *Cell*, 139, 234–244.

Chapter 3: Different Genes and Functions are Associated with the Offspring Size-Number Trade-off of Different Broods in *Daphnia*

The life history trade-off between offspring size and number of offspring, caused by differential allocation of resources required to maintain mass balance, represents a fundamental parent-offspring conflict. The molecular mechanisms underlying the trade-off remain largely unknown, however: to what extent is the trade-off genetically determined, and which genes are responsible for apportioning material among larger but fewer offspring, versus smaller and more numerous? I investigated the molecular basis of variation in the offspring size-number trade-off using phenotypic, gene expression, and SNP data from a well-studied *Daphnia* panel. I find (1) that different gene sets with different functional annotations are associated with subsequent broods; and (2) despite the phylogenetic breadth of the trade-off, the associated genes are neither evolutionarily conserved nor is their evolvability particularly low. The results suggest that the size-number allocation across broods shifts from bet-hedging to plasticity as mothers age, underlain (respectively) by core development and differentiation genes to metabolism-related genes.

INTRODUCTION

Trade-offs are a key aspect of life history theory: the trade-off between growth and reproduction is a classic example in which individuals can invest limited acquired resources in growth, which may increase future reproduction at the cost of current reproduction; or they may invest more resources in current reproduction at the cost of current growth. Beyond life history theory (Roff 1992, 2007), trade-offs are a central problem of multidimensional adaptation (de Mazancourt & Dieckmann 2004; Roff & Fairbairn 2007; Kirkpatrick 2009) and are a necessary condition of long-term ecological coexistence (Tilman 1982; Chase & Leibold 2003). While there are empirical examples of the absence of trade-offs when otherwise expected (e.g., Reznick et al. 2000), some may only be "apparent exceptions" of trait pairs where the trade-off actually spans three or more traits or where the genetics of acquisition and allocation hide a trade-off (Pease & Bull 1988; Houle 1991). In general, the presence and importance of life history trade-offs is well-supported by theory and data (Roff 1992 and references therein).

One well-studied life history trade-off occurs between offspring size and offspring number: parents can invest some finite amount of resources in either few, large or many, small offspring in proportions that maximize fitness (Smith & Fretwell 1974). The offspring size-number trade-off represents a fundamental parent-offspring conflict (Trivers 1974; Godfray 1995; Einum & Fleming 2000): parents should invest resources in proportions that maximizes their fitness, which may include sacrificing some offspring in the current brood for investment in future broods, while offspring fitness would be maximized by having parents invest all resources—including those required for the parent's continued survival—in the current brood. This conflict has important implications for understanding the evolution of trade-offs and the fate of alleles that act

disparately in offspring and parent (Godfray 1995; Kolm et al. 2006). A natural question is then, "What mechanisms act to appropriate resources to either many/small or few/large offspring and balance the parent-offspring conflict?" Specifically, to what extent is the trade-off genetically determined, how many of what kinds of genes underlie variation in the trade-off, and to what extent does environmental variation interact with genetic variation to strengthen, weaken, or otherwise alter the trade-off? In their review of the status of trade-off evolution research, Roff and Fairbairn (2007) emphasize that relatively little is known about the genetic mechanisms that maintain variation in life history trade-offs. Beyond the quantitative genetics of trade-offs, a recent review of correlated trait quantitative trait loci (QTL), primarily among plants, found that genotypic and QTL-effect correlations were largely concordant (Gardner and Latta 2007). The authors also found that overlapping QTL with opposite signs resulted in negatively correlated trait pairs approximately 2/3 of the time; the remaining cases appear to support Houle's (1991) proposal that loci affecting resource acquisition can generate positive correlations that swamp the negative correlation of resource allocation. Houle's (1991) model of the genetics of trade-offs between acquisition and allocation further indicated that the number of loci contributing to each component plays a critical role in the sign, so simply determining those numbers is a fundamental test of the model.

The common waterflea, *Daphnia pulex*, is a small freshwater crustacean that has served as a model ecological species for over a century; as a result I (collectively) have a detailed picture of the factors that influence its presence, abundance, and dynamics, as well as the species' ecological impacts (e.g., Lampert 2006). *Daphnia pulex* is a cyclical parthenogen in which all-female populations reproduce asexually for most of the growing season and releasing broods of neonates, and excepting obligate-parthenogen hybrids,

they undergo a bout of sexual reproduction at the end of the season to produce recombinant resting eggs. Mothers provision parthenogenetic neonates with resources in the form of lipid droplets that provide a buffer against starvation during the earliest instars, such that larger offspring possess higher fitness (Tessier & Goulden 1982; Tessier & Consolatti 1989). This provisioning is a basis for generating the offspring size-number trade-off, which was known from Cladocera as early as 1914 (Agar 1914) is often observed in *Daphnia*. The tradeoff has been found in several *D. magna* studies (e.g., Glazier 1992; Ebert 1993; Boersma 1997); Tessier and Consolatti (1991) found a strong negative relationship—and a significant effect of food level—in *D. pulex*; in his review of zooplankton quantitative genetics, Spitze (1995) reported a range of offspring size-number correlations from positive to negative, and in his own primary research experimentally found a positive relationship between offspring size and number in the presence of an invertebrate predator (Spitze 1991). In addition to our knowledge of the general presence and importance of the offspring size-number trade-off in *Daphnia*, the species' genome was recently sequenced (Colbourne et al. 2011) and an array of genomic and transcriptomic tools allow us to investigate the genes and loci underlying organismal variation. This knowledge of the importance of the offspring size-number trade-off in ecological and evolutionary contexts, combined with the genomic resources available for the species, makes *Daphnia* an ideal study system for investigating the problem.

Here I report on the quantitative genetics, transcriptomics, and molecular functions of genes associated with variation in the offspring size-number trade-off in *Daphnia*. I have previously described the phenotypic and genetic correlations between offspring size and number in the study panel; the genes whose expression is associated with the traits individually; and identified several candidate SNPs associated with the

traits individually (Chapters 1 and 2). Beyond simply identifying genes whose expression is correlated with variation of the offspring size-number trade-off, I test two hypotheses:

1. Different molecular mechanisms are responsible for the size-number trade-off in different broods, i.e., as mothers age.
2. The genes underlying the size-number trade-off are evolutionarily conserved.

The first hypothesis arises from my observation that the strength of the trade-off declines in this panel from the first ($r = -0.29$) to third brood ($r = 0.08$), and is strongest between offspring mass of the first brood and number of offspring in the second ($r = -0.38$). One explanation of the change is that the genetic basis of the relationship is overwhelmed by environmental effects as individuals age , and the low heritability of third brood mass ($H^2 = 0.477$) compared to first brood mass ($H^2 = 0.618$) suggests this may be the case. However, the third brood mean-scaled genetic variances (0.46 and 0.48 for mass and number, respectively) are only slightly lower than the values for the first brood (0.574 and 0.496, respectively) in the context of 32 measured traits (range = 0.13-3.85; Chapters 1-2). Two possibilities for explaining this pattern are that either the same genes increase in environmental sensitivity as mothers age, or that different genes are involved. On the basis that mothers have had longer to assess resource availability, and to assess prior reproductive success, I predict that there will be little overlap of the gene sets associated with the trade-off of the first and third broods, but that the genes associated with the first-brood mass to first- and second-brood number trade-offs are similar. Assuming selection for reproductive consistency early in life (bet-hedging; Kozlowski & Stearns 1989) and selection for ability to assess current conditions later in life to adjust allocation (i.e., plasticity), I predicted the first brood-associated genes will

be enriched for reproduction and development functional terms while the third brood-associated genes will be enriched for environment sensing and metabolic terms.

We pose Hypothesis 2 based on the observation that many taxa, from plants to humans, exhibit the offspring size-number trade-off (Roff 1992). Is it then possible that the genes associated with trade-off variation are evolutionarily conserved? First, I predict that trade-off-associated gene transcripts will possess lower evolvability—mean-scaled genetic variance in expression levels—than expected by chance. These values represent to the potential for selection to act on transcript abundance, and in theory, affect variation in the trade-off. Second, I predict then that the divergence of trade-off-associated proteins will be lower on average than the average divergence of all *Daphnia* genes, expressed genes, or differentially expressed genes. Furthermore, because *Daphnia* is a novel genomic model species and annotations are inferred from sequence similarity, then I expect a disproportionate number of genes associated with trade-off variation to be annotated in *Daphnia*.

RESULTS

While the offspring-size trade-offs are not particularly strong in the panel (genotypic values in Figure 3.1, phenotypic values in SI Figure 3.1), the common variance of each trade-off exhibited relatively high heritability, evolvability, and significant genetic effects for all but the third-brood trade-off (Table 3.1). After controlling for resource acquisition the strength of the third-brood offspring size-number correlation increased from 0.08 to 0.23 ($p = 0.18$). I have found non-linear relationships between gene expression and phenotypic variation using distance correlation in past work with this panel, and, despite the lack of a significant "clone" term, retained the third-

brood trade-off eigenvector for further analysis. There was substantial variation in inter-trade-off genetic correlations, ranging from negative values for slope variation and the brood 3 trade-off, to strongly positive between the common offspring variance and the first- and second-brood trade-offs (Table 3.2).

We identified numerous genes whose expression was strongly correlated with trade-off variation: there were between 117 and 274 genes associated with each trade-off trait (Table 3.3; SI Table 3.1). Overlap of the gene sets associated with each trade-off trait were as predicted from Hypothesis 1 and as expected given the inter-trait genetic correlations observed in the panel: different gene sets were associated with the first and third brood trade-offs, many genes were shared among the first-brood related trade-offs, and the gene sets associated with variation in size-number correlation were largely different from the common variance analyses (Table 3.4). Above the level of individual gene-trade-off relationships, I recovered interesting differences in the modularity of trade-off-specific gene sets. For example, variation in the second-brood number / first-brood mass correlation was associated with more transcripts than any other trait, but the genes are coexpressed in five tightly coexpressed modules (Figure 3.2A-B). In contrast, approximately the same number of transcripts were significantly associated with variation in the third-brood size-number trade-off, but clustered in twelve coexpression modules (Figure 3.2C-D). In addition to gene expression associated with trade-off variation, I identified 203 SNP-trait associations at $p < 0.01$, with six to 204 markers per trait (SI Table 3.2). As noted in Methods, FDR correction rendered all marker-trait association non-significant, and these represent markers significant at the nominal p-value.

The number of trade-off-associated genes possessing an annotation in each of the three GO domains was significantly higher than expected for most comparisons (Table

3.5). Many of the enriched GO terms matched expectations: first-brood trade-offs were enriched for genes with basic developmental and cell proliferation annotations, while metabolic and catabolic terms were enriched for the third-brood trade-off (SI Table 3.3). I summarize the enriched GO terms in Figure 3.3, which is developed in the Discussion.

Despite finding more annotated genes associated with trade-offs than expected—which suggests the gene sequences are more conserved than the average *Daphnia* gene—we did not find evidence of low transcript evolvability or amino acid similarity with Drosophila melanogaster. Mean evolvability of transcripts associated with most trade-off traits was not significantly different from mean DEG transcript evolvability ($p > 0.46$), except that the mean evolvability of transcripts associated with the correlation between first brood mass and second brood number of offspring (median = 0.927, s.e. = 0.034) was significantly higher than the mean DEG transcript evolvability (median = 0.844, s.e. = 0.007, $p = 0.0091$). Similarly, most measures of protein conservation relative to *D. melanogaster* proteins were not significantly different between gene sets of interest and reference sets (SI Table 3.4). The mean E-value of expressed genes, the gene set associated with variation in the slope of first-brood mass and second brood number, and the gene set associated with the first-brood PC were significantly lower than the mean E-value of all gene model matches (i.e., matches were stronger). However, the relative magnitude of the differences was very small and not accompanied by significant differences in metrics characterizing the divergence. Genes associated with the third-brood size-number trade-off possessed more Drosophila hits per *Daphnia* gene (x = 4.3, s.e. = 0.302) than the mean among DEGs (x = 3.62, s.e. = 0.055); in contrast, genes associated with the second-brood number / first-brood mass trade-off possessed fewer Drosophila hits per *Daphnia* gene (x = 2.8, s.e. = 0.27). I found, however, the statistical

significance of these patterns disappeared when I re-ran the analysis with a more stringent match threshold ($1e^{-25}$).

DISCUSSION

The relationship between the number and size of offspring produced in a given brood is a fundamental trade-off driven by limited resources and represents a classic parent-offspring conflict. Identifying the genetic variants underlying the offspring size-number trade-off will provide insight into the proximate mechanisms of variation and permit examination of the molecular evolution of this important aspect of life history. Here I have recovered a genetic component to trade-off variation in *Daphnia*, identified a host of candidate genes whose expression is strongly correlated with variation in the offspring size-number trade-off, and recovered support for the hypothesis that different gene sets, with distinct functional annotations, are associated with the trade-off across broods. There was, however, no support for the hypothesis that trade-off-associated genes are relatively evolutionarily conserved, nor that their transcripts possess lower evolutionary potential (i.e., evolvability).

I found that variation around most trade-offs contained a significant genetic component even though the negative correlations between offspring size and number in the *Daphnia* panel weren't extremely strong. This keeps with findings in several other systems, including other *Daphnia* populations (Lynch 1980; Spitze et al. 1991; Lynch et al. 1998; Dudycha & Tessier 1999), bank voles (Mappes & Koskela 2004), and clonal plants (Stuefer et al. 2002), where heritable variation for the trade-off has been recovered. By the third brood, when the size-number trade-off was non-existent, I observed genotypes such as Spitze's (Spitze 1991) "superfleas" (Reznick et al. 1998): clones in the

upper-right quadrant of Figure 3.1 are producing atypically many and large offspring. However, as Reznick and colleagues note, optimal life history trait values are context-specific. In the case of *Daphnia* offspring size and number both predation pressure (larger offspring are less susceptible to invertebrate predators) and resource availability (larger *Daphnia* may not be able to effectively exploit high resource concentrations; Tessier & Goulden 1982) dictate that large or small offspring are not always optimal, but depend on local conditions that may change rapidly across a heterogeneous landscape. An outstanding question is how this genetic variation might be maintained in natural populations. Roff and Fairbairn (2007) highlight spatio-temporal heterogeneity as a possible mechanism, while Lynch's and colleagues' (1998) work with *Daphnia pulex* mutation-accumulation lines suggests populations are at or near mutation-selection balance. These two options are not incompatible: gene flow between environmentally very different ponds, as is characteristic of the heterogeneous landscapes inhabited by *Daphnia*, is equivalent to mutation. The panel was collected from the range of habitats available to *Daphnia* in Michigan, and both landscape heterogeneity and mutation-selection balance likely explain a substantial proportion of the observed genetic variation.

Concomitant with the recovery of significant heritable variation from the quantitative genetic analysis, I found hundreds of genes whose expression was strongly correlated with different trade-off traits: SI Tables 3.1 and 3.2 provide numerous candidates to be subjected to additional testing. To our knowledge these constitute the first candidate genes specific to the offspring size-number trade-off, and are the first for *Daphnia*; as a consequence, there are no references against which I can compare these results to evaluate the consistency of results within or across species. Above the level of expression variation of individual genes, I recovered systematic variation in the topology

of coexpression networks of genes associated with each trait. Variation in coexpression network topologies is a common feature of transcriptomics research (this dissertation; Ayroles et al. 2009; Jumbo-Lucioni et al. 2010), but the biological and evolutionary importance of such variation is unclear (Hansen 2003; Espinosa-Soto & Wagner 2010); one possibility is that modularity variation reflects differences in trait evolvability (Wagner 1996; Hansen 2003; Hahn et al. 2004). I recovered a negative correlation for the four traits with evolvability estimates in this study, which, although the sample size is small, suggests a potentially important but unanswered question: are the organismal traits with the highest evolvability also those underlain by the fewest coexpression modules?

For the present purpose of insight into the molecular functions of the genes associated with the trade-off, examining the biological term enrichment analyses of the gene sets is more important than gene-wise or coexpression analysis. Analysis of the Gene Ontology terms of trade-off-associated genes suggests a testable model of the molecular functions underlying variation in the offspring size-number trade-off as mothers age: genes associated with variation in the first brood trade-off are reproduction- and development-related, but shift to metabolism and catabolism GO terms later in life (Figure 3.3). In addition, the model captures the frameworks of the "bet-hedging hypothesis" and a capacity for trade-off plasticity. First, the bet-hedging hypothesis posits that selection will act to maximize the geometric mean fitness rather than the arithmetic mean, and therefore mothers produce "some definite number" of offspring (Kozlowski and Stearns 1989). Consistency in number of the first brood driven by variation in core development and differentiation genes would guard against variation in resources across generations reflect a lineage's long-term geometric mean fitness; offspring size would follow from offspring number as a consequence, resulting in a

negative correlation between number and size. In contrast, as the mother ages and has assessed the environment for a longer period of time, allocation may be influenced more strongly by the available resources and the metabolic pathways required for processing those resources. This approach would afford greater plasticity in resource allocation as conditions change between broods. Furthermore, *Daphnia* and a variety of other species can re-absorb eggs or developing embryos (Kozlowski & Stearns 1989; Ohgushi 1996; Rosenheim et al. 2000), emphasizing a role for metabolic and catabolic processes that can reduce or remove the offspring size-number trade-off (Stibor 2002). The relatively short life-span of *Daphnia* in nature—less that 20d (from the life table data of Leibold 199x)—suggests that any broods beyond the first two would be "bonus" broods, and increased plasticity in allocation may be beneficial.

The lack of conservation of protein sequence and average (or in one case, slightly high) transcript evolvability for trade-off-associated genes was somewhat surprising, especially given that I found a disproportionate number of trade-off-associated genes possessed functional annotation. The generality of the trade-off suggested that the genes associated with the trait could be more strongly conserved than average, but that was not the case. One possible explanation is that my analysis was too simplistic and that more refined analysis would uncover a pattern in the evolutionary history or potential of genes associated with the trade-off. Additional work in this area would help resolve this question.

One limitation of this analysis is that the expression data was collected at the time of the first brood, not the third brood. During this interbrood period (x = 6.8d; Chapter 1) substantial expression changes may occur that could alter inferences (see, e.g., Hodgins-Davis & Townsend 2009). However, the proposed mechanism—a shift from genetic

control of development to terms associated with resource availability—and the recovery of metabolism-associated GO terms supports the hypothesis that different genes are associated with the trade-off as individuals mature. An alternative is that variation in third-brood trade-off was correlated with reproductive status at the time of tissue collection for gene expression analysis. While I took every precaution while collecting tissue for RNAseq (see Chapters 1 and 2), this correlation is possible. Future work with multiple expression assays through development, across a range of genotypes, and coupled with trade-off assays will certainly help refine our collective understanding.

Beyond identifying numerous candidate genes, I have presented a model relating two frameworks—bet-hedging and phenotypic plasticity—to biological functions of the genes associated with the offspring size-fecundity trade-off in *Daphnia*. Future work with a variety of species that exhibit the trade-off, for which there are a variety of experimental tools (e.g., QTL mapping, transgenic tools, etc.), will allow us to test the generality of the model and potentially link molecular functions to broader frameworks that organize our understanding of biological diversity.

METHODS

Panel description

Details of the panel and the methods used to collect phenotypic, expression, and genome-wide genetic variation have been described at length in Chapters 1 and 2. In brief, *Daphnia* clones were collected in Michigan, USA, from lake, shaded-pond, and sunny-pond habitats—capturing extensive phenotypic and genetic variation—near Kellogg Biological Station, and 32 clonal lines were established from single females.

The offspring data was collected in a growth and fecundity assay in which three replicates of each clone were raised for three generations (to minimize maternal effects) before . The first and third broods were counted, collected, desiccated, and weighed to the nearest 0.1μg on a Sartorius ultramicrobalance. Correlations of trait data were calculated using Spearman's rank correlation on genotypic values (i.e., trait means) because several relationships were non-linear. Gene expression was measured by 3'-tag RNAseq (Meyer et al. 2011) of three replicates per clone raised in the common-garden conditions of the phenotypic assays. Differential expression was determined using a generalized linear model with quasipoisson errors and variance inflation estimated for each gene from across all samples. In addition, each clone was genotyped at x = 3.3 million loci using RNAseq data and 2b-RAD genotyping (Wang et al. 2012) with the GATK (McKenna et al. 2010; DePristo et al. 2011).

Trade-off analyses

I considered two sets of traits in these analyses. The first set consists of common variance between pairs of traits—number and mass of the individuals in the first brood, number and mass of the individuals in the third brood, number of second brood and mass of the first brood individuals—represented by the first eigenvector. The PCA of each trade-off trait pair was calculated using the FactoMineR package of R 2.15.3 (R Development Core Team 2009) using genotypic (breeding) trait values, i.e., clone means from an ANOVA of form trait ~ clone. I also examine the first eigenvector (the only significant principal component by bootstrap resampling (Peres-Neto et al. 2005), $p < 0.001$) of the combined first-, second-, and third-brood data. The second trait set consists of clone-wise trade-off estimates: I calculated the within-clone correlation of

mass:number for the first and third broods, as well as first-brood mass on second-brood number, from the three replicates. Together these two sets allow us to consider variation around the trade-off and variation of the strength of the trade-off, respectively. As discussed in Results, the third-brood offspring mass-number correlation estimate is non-significant and slightly positive; because allocation trade-offs can be masked by acquisition correlations (Houle 1991), I tested if the third-brood correlation was influenced by resource acquisition variation by correlating the residuals of the linear model, [third-brood size, number] ~ growth rate.

To identify genes associated with the trade-offs, I calculated the distance correlation (Székely et al. 2007; Székely & Rizzo 2013) between clone means of 6434 differentially expressed genes (DEGs) and the trade-offs (eigenvectors or correlation estimates), and retained genes with Benjamini-Hochberg false discovery rate-corrected p-values < 0.05. I were interested in the overlap between gene sets associated with each trait, but a simple Fisher's Exact Test does not account for the non-independence of gene expression. Therefore, I evaluated the statistical significance of overlap by calculating the pairwise intersection of relevant trade-off contrasts using a randomization test that accounted for the non-independence of gene expression using the organization of global coexpression modules recovered by Malcom et al. (in prep., SI Code 1). After trade-off-associated transcripts were identified, I calculated the intersection with the gene sets associated with each main set (e.g., for the first-brood, the intersection of trade-off-associated genes with the union of offspring number and size gene sets). Last, I examined patterns of expression modularity of genes associated with each trade-off using the WGCNA package (Langfelder & Horvath 2008) for R 2.15.3.

To identify SNPs whose variants are predictive of the trade-offs, I performed a genome wide association analysis of 3449 markers against trade-off data using a linear model. FDR correction rendered all results non-significant ($p_{adjusted} \geq 0.17$), but because this is exploratory and little is known about the genetics of the offspring size-number trade-off in *Daphnia*, I report marker-trait associations that were significant at a nominal $p < 0.01$. After determining the genes within 1kb of each significant SNP, I performed Gene Set Enrichment Analyses (GSEAs) on the Gene Ontology (GO) terms of each set of significantly trait-associated genes and SNP-associated gene set using the GOstats package for R (Falcon & Gentleman 2011).

We evaluated the evolutionary conservation of trade-off-associated genes in terms of expression evolvability—a function of cis- and trans-acting mutations—and protein divergence. For transcript evolvability I calculated the mean-scaled genetic variance of all transcripts, DEG transcripts, and trade-off-associated transcripts. I quantified protein divergence in terms of the median E-value, mismatch and gap rate (per amino acid), and percent identity using BLASTP (Altschul et al. 1997) of *Daphnia* amino acids against the Drosophila melanogaster amino acid database. Evolvability was compared to the evolvability of all DEGs (whose evolvability is significantly higher than non-DEGs; Malcom et al. *in prep*.) by t-test of log-transformed data, and I examined the 95% confidence intervals of protein divergence summaries to determine if any large and perhaps biologically important differences were evident. Note that in neither case am I attempting to identify the genetic cause of evolvability or protein conservation differences (i.e., the responsible sequence variants), but rather to simply quantify the differences in two molecular phenotypes.

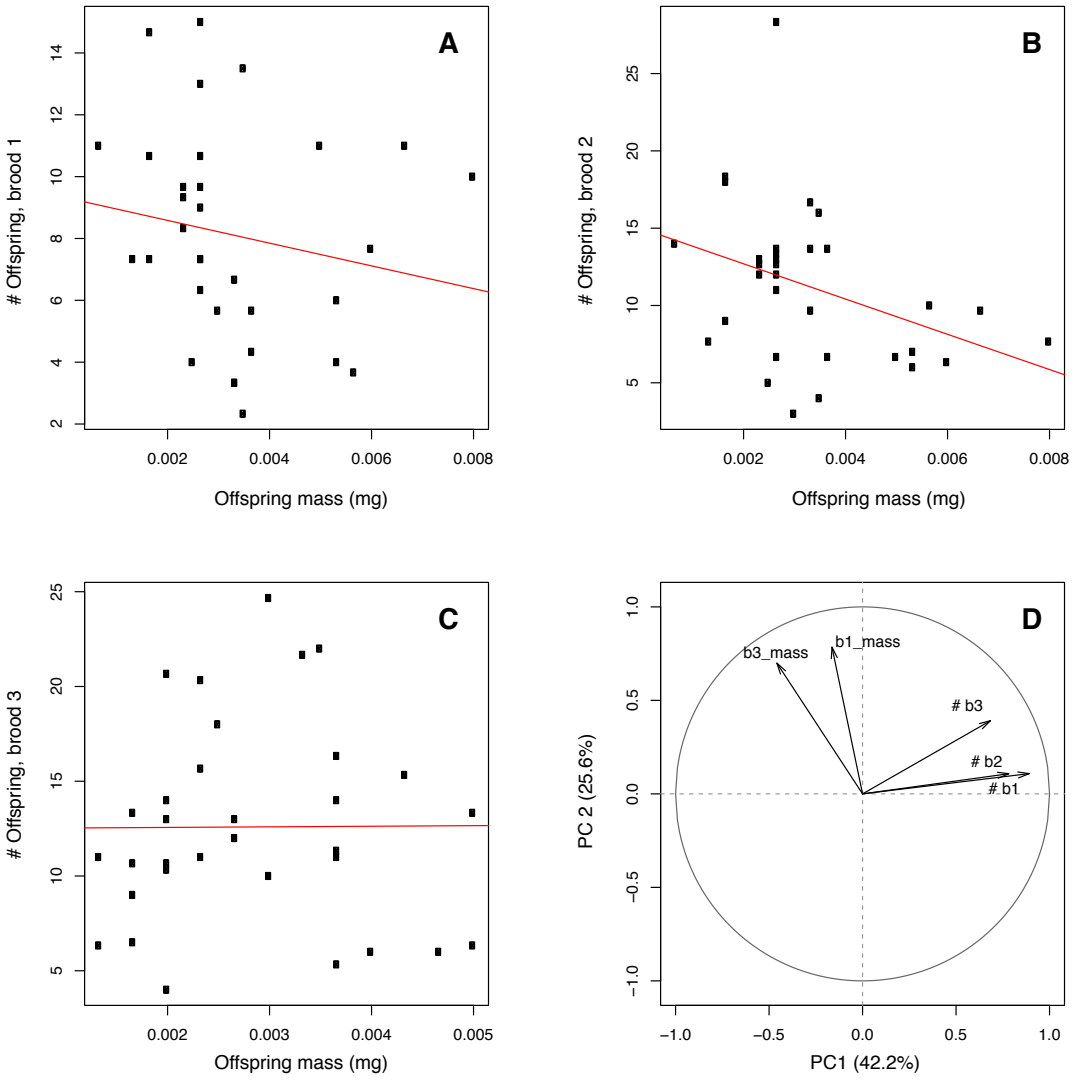


Figure 3.1. Offspring size-number trade-offs of genotypic trait values in the *Daphnia* panel. First-brood offspring mass and number are negatively correlated (r = -0.28, $p = 0.05$; panel A); the second-brood offspring number is more-strongly negatively correlated with first-brood offspring mass (r = -0.39, $p = 0.02$; panel B); but the third-brood trade-off is non-existent to slightly positive (r = 0.08, $p > 0.1$; panel C). The first two axes of variation among the offspring traits highlight the loadings and relationships of the five main traits from which trade-offs are derived.

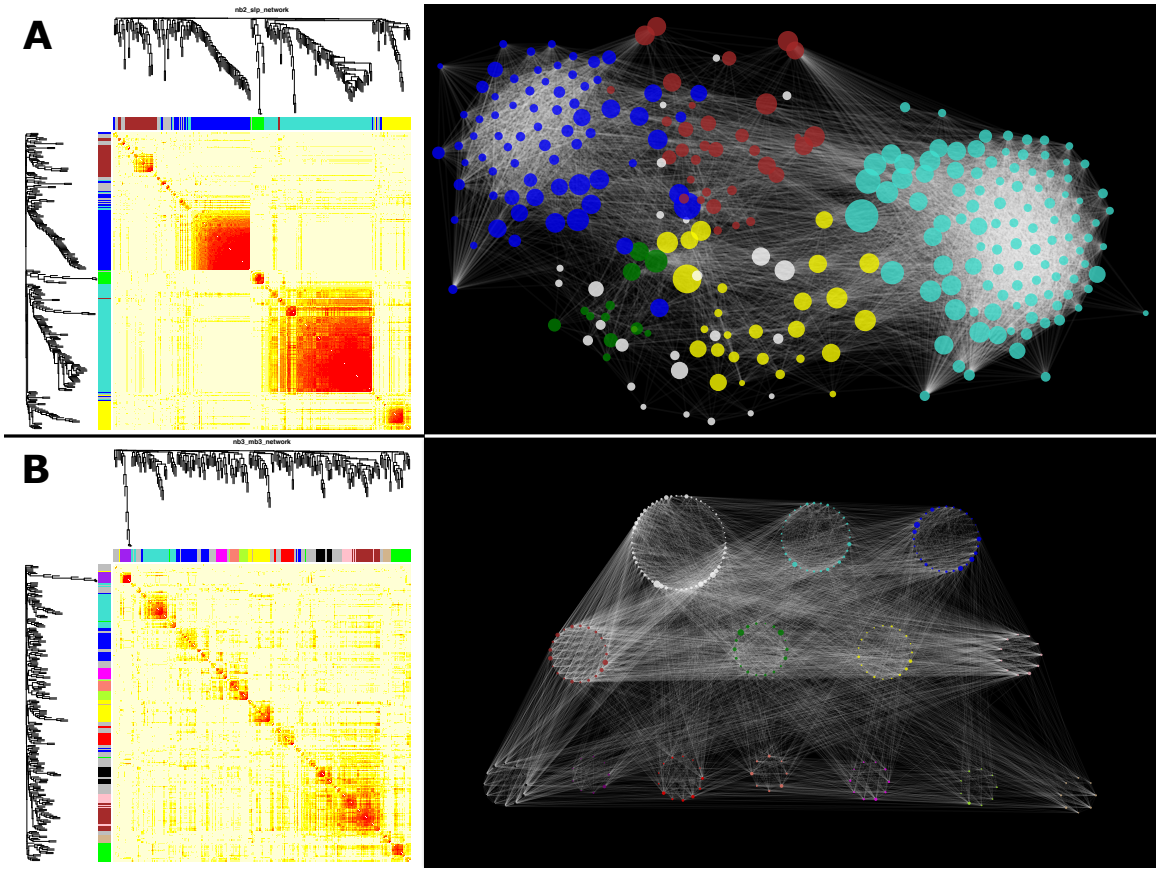


Figure 3.2. ***Daphnia*** **gene coexpression networks associated with (A) variation in the first-brood mass, second-brood number correlation and (B) variation in the third-brood size-number trade-off.** The left sub-panel of each part is a heatmap of coexpression modules (from WGCNA), and each is translated to the respective network view in the right sub-panel. Network layouts differ to highlight the structural differences. Node size is proportional to betweenness-centrality (i.e., the importance of a node in shortest paths between node pairs), and edge weight is proportional to the strength of the correlation between nodes. While the number of genes is similar, modularity is very different between the networks associated with the traits.

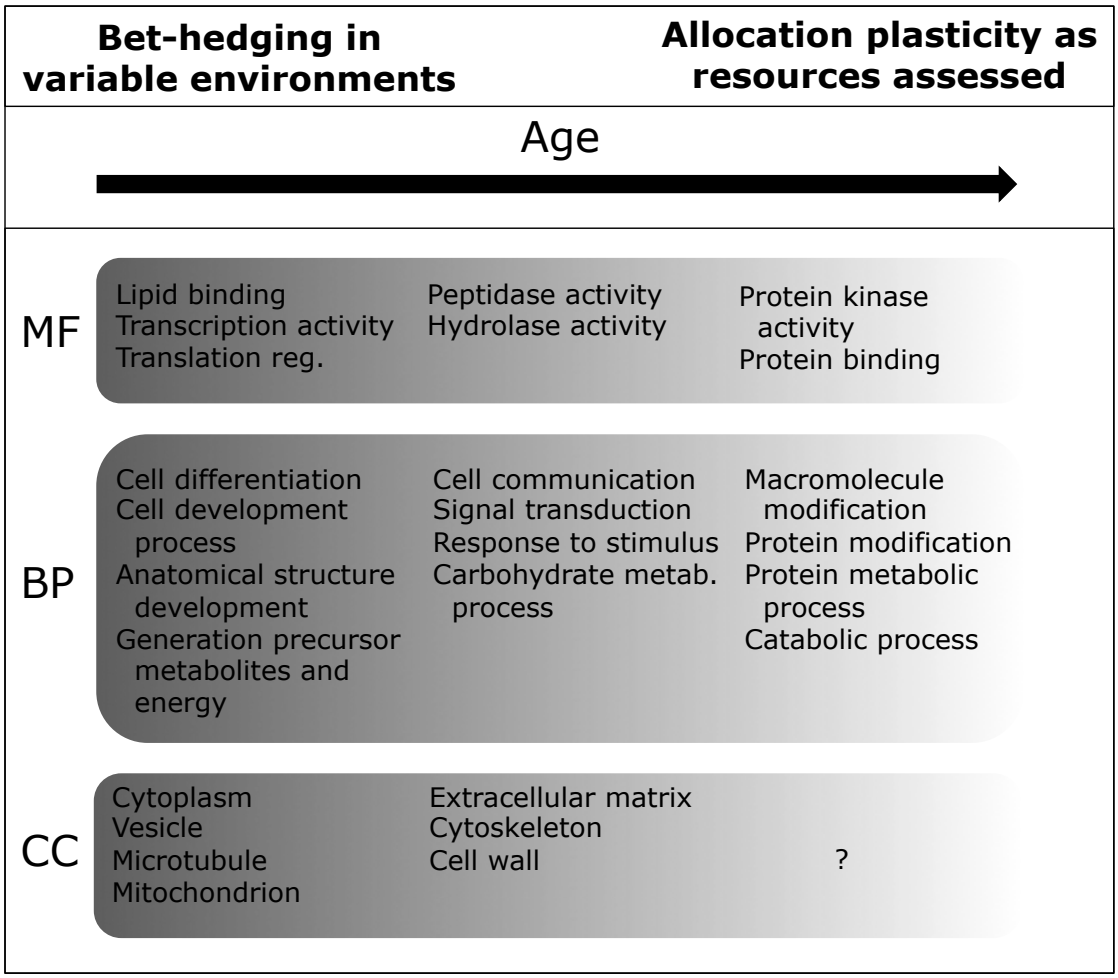


Figure 3.3. A model of the frameworks of offspring size-number allocation (top) and associated biological functions (bottom) as an iteroparous individual ages. In bet-hedging, geometric mean fitness is selected for and variation in core development and differentiation genes regulate the trade-off. In subsequent broods, as the mother has had longer to assess the environment, plasticity is the driving framework and metabolism-associated genes regulate the trade-off (if any). Biological functions are grouped according to Gene Ontology domains, MF = Molecular Function; BP = Biological Process; CC = Cellular Component.

Tradeoff	H^2	V_P	V_G	I_G	genotype *p*
n_b1-m_b1	0.624	1.178	0.735	0.857	0.0438
n_b3-m_b3	0.514	1.173	0.603	0.777	0.415
n_b2-m_b1	0.636	1.243	0.791	0.889	0.03
offsp. Com. Var	0.655	1.882	1.233	1.11	0.0155

Table 3.1. Quantitative genetics of *Daphnia* offspring size-number trade-off. "genotype *p*" is the *p*-value of the "clone" term in the linear model trade-off ~ clone.

	n_b1-m_b1	n_b3-m_b3	n_b2-m_b1	Offsp. Com. Var	brood 1 slope	brood 2 slope
n_b1-m_b1	1					
n_b3-m_b3	0.028	1				
n_b2-m_b1	0.827	0.135	1			
offsp. Com. Var	0.907	0.275	0.892	1		
brood 1 slope	0.158	-0.132	0.012	0.03	1	
brood 2 slope*	0.107	-0.11	0.147	0.027	0.212	1

Table 3.2. Genetic correlations between *Daphnia* offspring size-number trade-offs. * "Brood 2 slope" is the slope of the first-brood mass, second-brood number trade-off.

Tradeoff	# genes	# modules	T1 # genes	T2 # genes	total # overlap	total % overlap
n_b1-m_b1	119	6	132	140	41	15.1
n_b3-m_b3	256	13	160	261	91	22.5
n_b2-m_b1	146	6	196	140	71	22
offsp. Com. Var	129	4	--	--	--	--
brood 1 slope	117	7	132	140	3	1.1
brood 2 slope*	274	5	196	140	11	3.5

Table 3.3. Summary statistics of transcript-trade-off associations. For comparison I provide the number of genes associated with the main traits (T1 = first trait of trade-off, T2 = second trait of trade-off) and the overlap with trade-off-associated genes.

	n_b1-m_b1	n_b3-m_b3	n_b2-m_b1	offsp. Com. Var	brood 1 slope	brood 2 slope
n_b3-m_b3	2 (0.777)	--				
n_b2-m_b1	42 (< 0.001)	6 (0.335)	--			
offsp. Com. Var	60 (< 0.001)	9 (0.011)	71 (< 0.001)	--		
brood 1 slope	4 (0.043)	9 (0.038)	1 (0.701)	0 (0.865)	--	
brood 2 slope	3 (0.956)	6 (0.608)	3 (0.792)	3 (0.989)	4 (0.318)	--

Table 3.4. Observed, expected, and p-values of overlap between trade-off gene sets.

Trait	# genes	# MF ann. (proportion)	# BP ann. (proportion)	# CC ann. (proportion)	MF expected	BP expected	CC expected	MF Chi^2	BP Chi^2	CC Chi^2	MF *p*-value	BP *p*-value	CC *p*-value
n_b1-m_b1	120	36 (0.30)	42 (0.35)	39 (0.33)	28.48	32.24	29.55	1.99	2.95	3.02	1.59E-01	8.57E-02	8.22E-02
offsp. Com. Var	129	43 (0.33)	52 (0.40)	47 (0.36)	30.61	34.66	31.77	5.01	8.68	7.30	2.51E-02	3.22E-03	6.89E-03
brood 1 slope	117	35 (0.30)	42 (0.36)	37 (0.32)	27.76	31.43	28.81	1.89	3.55	2.33	1.70E-01	5.95E-02	1.27E-01
n_b2-m_b1	146	48 (0.33)	61 (0.42)	56 (0.38)	34.65	39.23	35.96	5.15	12.09	11.17	2.33E-02	5.08E-04	8.29E-04
brood 2 slope*	274	95 (0.35)	106 (0.39)	95 (0.35)	65.02	73.62	67.48	13.82	14.24	11.23	2.01E-04	1.60E-04	8.07E-04
n_b3-m_b3	256	88 (0.34)	98 (0.38)	99 (0.39)	60.75	68.78	63.04	12.22	12.41	20.51	4.72E-04	4.26E-04	5.95E-06

Table 3.5. Gene Ontology (GO) term enrichment summary by trade-off trait and GO domain. * MF = molecular function; BP = biological process; CC = cellular component ** Proportion of *Daphnia* genes with annotations: MF = 0.237; BP = 0.269; CC = 0.246

LITERATURE CITED

Agar, W.E. (1914). Experiments on Inheritance in Parthenogenesis. *Philos. Trans. R. Soc. Lond. Ser. B Contain. Pap. Biol. Character*, 205, 421–489.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., *et al.* (2009). Systems genetics of complex traits in Drosophila melanogaster. *Nat Genet*, 41, 299–307.

Boersma, M. (1997). Offspring size and parental fitness in *Daphnia magna*. *Evol. Ecol.*, 11, 439–450.

Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University Of Chicago Press.

Colbourne, J.K., Pfrender, M.E., Gilbert, *D.*, Thomas, W.K., Tucker, A., Oakley, T.H., *et al.* (2011). The Ecoresponsive Genome of *Daphnia pulex*. *Science*, 331, 555 – 561.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, 43, 491–498.

Dudycha, J.L. & Tessier, A.J. (1999). Natural Genetic Variation of Life Span, Reproduction, and Juvenile Growth in *Daphnia*. *Evolution*, 53, 1744–1756.

Ebert, *D.* (1993). The trade-off between offspring size and number in *Daphnia magna*: the influence of genetic, environmental and maternal effects. *Arch. Hydrobiol.-Suppl.-*, 90, 453–453.

Einum, S. & Fleming, I.A. (2000). Highly fecund mothers sacrifice offspring survival to maximize fitness. *Nature*, 405, 565–567.

Espinosa-Soto, C. & Wagner, A. (2010). Specialization Can Drive the Evolution of Modularity. *Plos Comput Biol*, 6, e1000719.

Falcon, S. & Gentleman, R. (2011). How To Use GOstats Testing Gene Lists for GO Term Association.

Glazier, *D.*S. (1992). Effects of Food, Genotype, and Maternal Size and Age on Offspring Investment in *Daphnia Magna*. *Ecology*, 73, 910–926.

Godfray, H.C.J. (1995). Evolutionary theory of parent offspring conflict. *Nature,* 376, 133-138.

Hahn, M., Conant, G. & Wagner, A. (2004). Molecular Evolution in Large Genetic Networks: Does Connectivity Equal Constraint? *J. Mol. Evol.*, 58, 203–211.

Hansen, T.F. (2003). Is modularity necessary for evolvability?: Remarks on the relationship between pleiotropy and evolvability. *Biosystems*, 69, 83–94.

Hodgins-Davis, A. & Townsend, J.P. (2009). Evolving gene expression: from G to E to G× E. *Trends Ecol. Evol.*, 24, 649–658.

Houle, *D.* (1991). Genetic Covariance of Fitness Correlates: What Genetic Correlations are Made of and Why it Matters. *Evolution*, 45, 630–648.

Jumbo-Lucioni, P., Ayroles, J., Chambers, M., Jordan, K., Leips, J., Mackay, T., *et al.* (2010). Systems genetics analysis of body weight and energy metabolism traits in Drosophila melanogaster. *Bmc Genomics*, 11, 297.

Kirkpatrick, M. (2009). Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, 136, 271–284.

Kolm, N., Goodwin, N.B., Balshine, S. & Reynolds, J.D. (2006). Life history evolution in cichlids 2: directional evolution of the trade-off between egg number and egg size. *J. Evol. Biol.*, 19, 76–84.

Kozlowski, J. & Stearns, S.C. (1989). Hypotheses for the Production of Excess Zygotes: Models of Bet-Hedging and Selective Abortion. *Evolution*, 43, 1369–1377.

Lampert, W. (2006). *Daphnia*: model herbivore, predator and prey. *Pol. J. Ecol.*, 54, 607–620.

Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.

Lynch, M. (1980). The Evolution of Cladoceran Life Histories. *Q. Rev. Biol.*, 55, 23–42.

Lynch, M., Latta, L., Hicks, J. & Giorgianni, M. (1998). Mutation, Selection, and the Maintenance of Life-History Variation in a Natural Population. *Evolution*, 52, 727–733.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Candidate loci, genes, and networks underlying ecologically important behavioral variation in a model ecological species, *Daphnia pulex*.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Exploring the molecular basis of variation in size and fecundity in an emerging model organism, *Daphnia pulex*.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). The joys of emerging model genomic species: Evolutionary patterns and molecular biological function in *Daphnia*.

Mappes, T. & Koskela, E. (2004). Genetic Basis of the Trade-Off Between Offspring Number and Quality in the Bank Vole. *Evolution*, 58, 645–650.

De Mazancourt, C. & Dieckmann, U. (2004). Trade-Off Geometries and Frequency-Dependent Selection. *Am. Nat.*, 164, 765–778.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20, 1297–1303.

Meyer, E., Aglyamova, G.V. & Matz, M.V. (2011). Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol. Ecol.*, 20, 3599–3616.

Ohgushi, T. (1996). A reproductive tradeoff in an herbivorous lady beetle: egg resorption and female survival. *Oecologia*, 106, 345–351.

Pease, C.M. & Bull, J.J. (1988). A critique of methods for measuring life history trade-offs. *J. Evol. Biol.*, 1, 293–303.

Peres-Neto, P.R., Jackson, *D.*A. & Somers, K.M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Comput. Stat. Data Anal.*, 49, 974–997.

R Development Core Team. (2009). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Reznick, *D.*, Nunney, L. & Tessier, A. (2000). Big houses, big cars, superfleas and the costs of reproduction. *Trends Ecol. Evol.*, 15, 421–425.

Roff, *D.*A. (1992). *The evolution of life histories: theory and analysis*. Springer.

Roff, *D.*A. (2007). Contributions of genomics to life-history theory. *Nat Rev Genet*, 8, 116–25.

Roff, *D.*A. & Fairbairn, *D.*J. (2007). The evolution of trade-offs: where are we? *J. Evol. Biol.*, 20, 433–447.

Rosenheim, J.A., Heimpel, G.E. & Mangel, M. (2000). Egg maturation, egg resorption and the costliness of transient egg limitation in insects. *Proc. R. Soc. Lond. B Biol. Sci.*, 267, 1565–1573.

Smith, C.C. & Fretwell, S.D. (1974). The Optimal Balance between Size and Number of Offspring. *Am. Nat.*, 108, 499–506.

Spitze, K. (1991). *Chaoborus* Predation and Life-History Evolution in *Daphnia pulex*: Temporal Pattern of Population Diversity, Fitness, and Mean Life History. *Evolution*, 45, 82–92.

Spitze, K. (1995). Quantitative genetics of zooplankton life histories. *Experientia*, 51, 454–464.

Spitze, K., Burnson, J. & Lynch, M. (1991). The covariance structure of life-history characters in *Daphnia pulex*. *Evolution*, 45, 1081–1090.

Stibor, H. (2002). The role of yolk protein dynamics and predator kairomones for the life history of *Daphnia magna*. *Ecology*, 83, 362–369.

Stuefer, J.F., Van Hulzen, J. b. & During, H.J. (2002). A genotypic trade-off between the number and size of clonal offspring in the stoloniferous herb Potentilla reptans. *J. Evol. Biol.*, 15, 880–884.

Székely, G.J. & Rizzo, M.L. (2013). The distance correlation -test of independence in high dimension. *J. Multivar. Anal.*, 117, 193–213.

Székely, G.J., Rizzo, M.L. & Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Stat.*, 35, 2769–2794.

Tessier, A. & Goulden, C. (1982). Estimating food limitation in cladoceran populations. *Limnol. Ocean.*, 27, 707–717.

Tessier, A.J. & Consolatti, N.L. (1989). Variation in Offspring Size in *Daphnia* and Consequences for Individual Fitness. *Oikos*, 56, 269–276.

Tessier, A.J. & Consolatti, N.L. (1991). Resource Quantity and Offspring Quality in *Daphnia*. *Ecology*, 72, 468–478.

Tilman, *D.* (1982). *Resource competition and community structure*. Princeton University Press.

Trivers, R.L. (1974). Parent-offspring conflict. *Am. Zool.*, 14, 249–264.

Wagner, A. (1996). Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biol. Cybern.*, 74, 557–567.

Wang, S., Meyer, E., McKay, J.K. & Matz, M.V. (2012). 2b-rad: a simple and flexible method for genome-wide genotyping. *Nat. Methods*, 9, 808–810.

Chapter 4: Genomic Analyses Suggest Two Competing Hypotheses for a "Keystone Gene Network" in an Ecological Keystone Species, *Daphnia pulex*

Keystone species are those whose effects on community composition and dynamics are greater than expected given their abundance. If we identify the genes responsible for variation in a keystone species' niche, then we extend our understanding of molecular variation beyond the organism to communities. We leveraged the well-known ecology of *Daphnia pulex*—a keystone species of freshwater ponds—and its recently developed genomic resources to identify two competing "keystone gene networks" associated with variation in three traits important to defining the species' niche. Quantitative genetic analysis of the trait variation suggests substantial shared genetic variation that should typically facilitate multidimensional adaptation given the common selective regimes *Daphnia* face. One candidate network associated with this variation includes 34 genes that are over-enriched for molecular and biological terms related to oxygen-handling and response to stimuli; the alternate network spans 135 genes that encompass a wide variety of molecular and biological terms. We also identify numerous promoter motifs and sequence variants unique to and shared among the candidate networks. The identification of two competing candidate keystone gene networks highlights the utility of genomic approaches for developing novel hypotheses, in this case, to advance our understanding of the heritable bases of niche-mediating variation in an ecological keystone species.

INTRODUCTION

Explaining the origin and maintenance of biodiversity can span many different levels of biological organization, from genes to communities. Genetic and environmental variation interact to shape trait variation; trait variation shapes the interactions between individuals and their environment; and variation in the biotic and abiotic environment shapes the genes that are present in future generations, thus linking the levels of organization. Despite our ability to articulate this heuristic overview, we have very few empirical examples of the mappings across each of these levels of organization because either (a) our knowledge of the ecology of model genetic species is lacking or (b) we lack genetic resources for model ecological species (Feder & Mitchell-Olds 2003; Simon et al. 2011).

A potentially powerful approach to address the question of the genes most-important to ecological communities could begin with a focus on keystone species, i.e., species whose presence and abundance dramatically affect community structure, dynamics, and stability (Paine 1969; Mills et al. 1993; Leibold 1996). By identifying the genetic bases of variation in the traits that shape a keystone species' niche (Chase & Leibold 2003), we extend the explanatory power of molecular variants beyond their effects on a single species' phenotype to the effects on communities. This effectively places genetic and molecular research in the context of Dawkins's (1982) "extended phenotype", and to an extent, "niche construction" (Odling-Smee et al. 2003). Furthermore, while focusing on the molecular basis of variation in individual traits may be informative for many questions (Mackay 2001; Mackay et al. 2009), genetic correlations among the niche-mediating traits may hinder or accelerate multidimensional adaptation (Lande 1979; Lande & Arnold 1983; Arnold et al. 2008). In the context of a

keystone species, mutations that simultaneously affect several niche-mediating traits could have a dramatic effect on eco-evolutionary dynamics of communities (Hairston et al. 2005; Carroll et al. 2007; Ellner et al. 2011). Our goal then should be to identify genes that simultaneously affect variation in multiple niche-mediating traits of keystone species. That is, if we can identify a species with a host of genetic resources and that our ecological knowledge indicates is a keystone species, we can begin to address the question, What genes and loci are most important for shaping ecological communities?

With the recent sequencing of the *Daphnia pulex* genome (Colbourne et al. 2011) we have an ideal system for connecting genetic, phenotypic, population, and community variation. Over a century's worth of ecological research has established the factors most important to the ecological requirements and impacts of *Daphnia* (Forbes 1880; Lampert 2006). They are keystone species because of their central role in ecological food webs, shaping algae and protist communities through grazing, serving as a major resource for vertebrate and invertebrate predators, and forming a key link for both spatial transfer of materials and energy transfer through food webs (Leibold 1989; Sarnelle 1993, 2005; Carpenter & Kitchell 1996; Persson et al. 2007). Furthermore, we have a good understanding of the traits that mediate the *Daphnia* niche in an variety of ecological scenarios: body size is central to mediating interactions with competitors and predators (Brooks & Dodson 1965; Goulden et al. 1982; Cernỳ & Bytel 1991; Tessier et al. 1992; Wolinska et al. 2007); phototactic behavior affects both predation risk and the quality and quantity of available food (De Meester 1991; Pijanowska et al. 1993; De Meester & Weider 1999; Hood & Sterner 2010); and swimming speed regulates food consumption rates, the probability of detection by predators, and the escape probability from invertebrate predators (Kerfoot et al. 1980; Dodson & Ramcharan 1991; Seuront et al.

2004; Boeing et al. 2006). Note that throughout we refer to "*Daphnia*" generically but are referring specifically to the *D. pulex* species group. The contorted status of the species group is discussed in CITES.

Our goal was to integrate the recently developed genomic resources for *Daphnia* with the extensive knowledge of the ecological factors that shape and are shaped by the species to identify a candidate ecological "keystone gene network". Genomic approaches to such problems are inherently about generating hypotheses (Biesecker 2013), and rather than identify a single candidate we performed two analyses that give different, but slightly overlapping, results. These inferences are based on a panel of 32 wild-collected *Daphnia* clones from southern Michigan, including the classic *D. pulex* ecotype that inhabit shaded, fishless ponds with abundant invertebrate predators; the classic *D. pulicaria* ecotype that inhabits lakes containing fish and co-migrating invertebrate predators; and hybrids between the two parental ecotypes, which inhabit sunny, fishless ponds; so that the range of ecological possibilities for the species was captured. We have previously described the genetic, gene expression, and variation in 32 organismal traits in this panel (Chapters 1 and 2): linkage disequilibrium among 6338 SNPs decays to background levels by ~200bp; there are 6434 differentially expressed genes (DEGs) under common-garden conditions, i.e., expression differences due only to genotypic difference between clones, which cluster into 27 global coexpression modules; and there is extensive variation in life history, swimming, and phototaxis traits. Our results both provide an extensive set of hypotheses for future experimental testing, and illustrate a promising approach to extending our understanding of genetic variation to the level of communities.

RESULTS

There was three- to eight-fold among-clone variation in the 32 measured traits, and in all cases, the trait values were continuously and normal or near-normal distributed (Fig. 1A-C). The first five principal components (PCs) among all genotypic trait values were statistically significant (Malcom et al. *in prep*.), and projecting the first two PCs illustrates the loadings of our three focal traits discussed below (Fig. 1D). Considering the genotypic variation among the three keystone traits, PC1 accounted for 52.9% of variance and was significantly correlated with all three traits (r^2 = 0.449, 0.515, 0.623 for swimming speed, growth rate, and phototaxis variation, respectively; SI Figure 4.1). The second PC, which is significantly associated with swimming speed and growth rate, accounts for half as much variation as PC1 (26.5%) and corresponds to an axis of variation orthogonal to remaining variation in phototaxis.

Candidate Network 1 (CN1) builds directly on the gene sets associated with growth rate, swimming speed, and phototaxis variation from our previous work with the panel (Chapters 1 and 2). Thirty-four genes formed a single, tight coexpression network (Figure 2A, SI Table 4.1) and all were from a single global coexpression module, the "purple" module of SI Figure 1.2. Candidate Network 2 (CN2) was identified as the network of genes whose coexpression was significantly associated with variation around PC1 (i.e., with the first eigenvector). This network encompassed 135 genes that clustered into seven distinct modules each with 7-26 tightly coexpressed genes per module (Figure 2B, SI Table 3.2), with genes drawn from 25 of the 27 global coexpression modules. Just three genes were present in both networks, two of which have roles in transcriptional regulation (Table 4.1).

Beyond identifying individual genes associated with trait variation we are often interested in understanding common patterns of the biological and molecular functions of these sets of genes. We found that the genes of CN1 were enriched for oxygen-handling, stimulus response, and "extracellular region" annotations (Table 4.2, top). In contrast, the genes of CN2 are enriched for catalytic, peptidase, and hydrolase activity Gene Ontology (GO) Molecular Functions; several metabolism-centered Biological Processes; and four "extracellular region" Cellular Component terms (Table 4.2, bottom). Individual coexpression modules of CN2 were enriched for specific sets of GO terms (SI Table 4.3). Enriched terms overlapping between the genes of the two candidate networks included "peptidase activity", "protein metabolic process", "extracellular region", and "antioxidant activity".

One mechanism driving gene coexpression is shared regulators, and we hypothesized that we would identify enriched promoter motif sequences that are targets of the regulators of the keystone network genes. From among the 34 CN1 genes we identified five promoter motifs enriched at $E < 0.05$ (Table 4.3, top); Motif 2 (GRGGGKGGKRGGRDRGG) was highly similar to several known promoters in *Drosophila*, with a top match ($q = 1.5e^{-6}$) to the transcription factor lethal (3) neo38, which is involved in chromatin silencing regulation and morphogenesis. From among CN2 genes we identified nine promoter motifs enriched at $E < 1e^{-3}$ (see Methods), for 9-65 genes per motif (Table 4.3, bottom). Three motifs were over-represented among the genes of two coexpression modules: seven genes in the "turquoise" module possess motifs 9 and 6 (expected = 2.9; $\chi^2 = 5.85$; $p = 0.0156$); and two genes of the "red" module possess motif 3 (expected = 0.47; $\chi^2 = 5.03$; $p = 0.0248$). Two of these motifs, Motif 1 and Motif 8, were highly similar ($q < 0.05$) to *Drosophila* motifs that are targets of

several transcription factors (Motif 1) and *jigr1* (Motif 8), the latter of which is associated with larval nervous system. Only a single motif was common between the genes of the two candidate networks, CN1 Motif 2 and CN2 Motif 1 ($q = 4.39e^{-6}$), both of which are G-rich and relatively unstructured.

DISCUSSION

Identifying the molecular basis of ecologically important trait variation in an ecological keystone species offers an opportunity to connect functional genetic variation in one species to variation in entire communities. Here we have investigated the model ecological keystone species *Daphnia pulex*, described the quantitative genetic covariances of three traits most-important to defining the species' niche, and identified two competing candidate "keystone gene networks" that may drive the heritable variation among these traits that are known to shape limnetic communities within and across trophic levels. One interpretation of our genomic analyses results is to treat the elements overlapping between analyses as "high-confidence" associations that are top candidates for follow-up investigation. However, assessing multiple competing hypotheses has long been considered good scientific practice (Chamberlin 1890), and although there is some overlap at each of the levels we examined, the candidate networks are largely distinct and offer distinct hypotheses about the genes and molecular functions that drive *Daphnia* "keystoneness".

The phenotypic variation we recovered in our panel was extensive, yet a substantial portion of genetic variation is shared among the three traits central to defining *Daphnia*'s role as a keystone species. This shared variation can be both a benefit and a hindrance for populations: if selection correlates with this variation, then populations will

adapt faster than if all traits are genetically independent (Kirkpatrick 2009; Lovell et al. 2013). By contrast, if the axes of selection are orthogonal (or near-orthogonal) to the axes of variation, then variation concomitant with one axis of selection may constrain adaptation along other axes. *Daphnia* have been models in investigations of rapid adaptation and its ecological consequences (Hairston et al. 1999, 2005; Miner et al. 2012), and our result that over half the genetic variation among the three keystone traits is shared along the first PC suggests high potential for rapid adaptation given the contrasting environments characteristic of the landscapes in which *Daphnia* commonly live. Clones that are smaller, slower, and possess a stronger response to fish kairomones should have higher fitness in large, fish-occupied lakes. In contrast, larger and faster-swimming clones that do not require strong fish kairomone response should have higher fitness in shallow, typically temporary ponds lacking fish. If *Daphnia* encounter conditions that require both strong kairomone response and fast growth or swimming, however, the shared variance will likely constrain the response to selection. This latter scenario should be rare, and the quantitative genetic results we recovered correspond to the major axis of genetic variation; they reinforce the idea that there is sufficient shared heritable variation to facilitate rapid adaptation in complex natural systems. Furthermore, this is an important result with respect to the Monopolization Hypothesis, which posits that strong population structure across a landscape in the face of high dispersal arises from rapid colonization and adaptation to local conditions (De Meester et al. 2002), and which was developed with zooplankton such as *Daphnia* in mind. The shared genetic variation among traits that mediate *Daphnia*'s niche should be sufficient for the monopolization process to operate efficiently at the landscape level. Note that our

analyses are based total genetic variance, not additive variance, and are appropriate for predicting within-season response to selection rather than between-season response.

In light of the eco-evolutionary relevance of our quantitative genetics results, identifying gene networks underlying the genetic (co-)variation holds the potential to elegantly extend our understanding of the consequences of molecular variation to the level of ecological communities. Our candidate networks suggest two rather different outlooks on the complexity of the molecular basis of *Daphnia* "keystoneness": CN1 suggests a relatively small number of genes mediating oxygen handling and response to external stimuli are responsible, while CN2 suggests a much more complex system, both in terms of number of genes and variety of biological processes. Our approaches to defining each of the two candidate networks are very different, and CN1 might be considered by some to be too naïve of an analysis while CN2 is statistically more appropriate. However, the biological interpretation of CN1 results is arguably more clear. The lentic systems inhabited by *Daphnia* are characterized by wide variation in oxygenation (Wetzel 2001), and variation in growth, swimming speed, and phototactic behavior should all depend on oxygen handling. Furthermore, the role of vertebrate and invertebrate predators in shaping *Daphnia* populations is well-known (Leibold & Tessier 1991; Spitze 1991; Dini & Carpenter 1992; Gliwicz & Umana 1994; Lampert 2006; Scoville & Pfrender 2010), so the enrichment of genes with "response to external stimuli" is a sensible result.

In contrast to CN1, CN2 contains four times as many genes, and among these there is an over-representation of genes annotated with development- and metabolism-associated terms. We know for *Daphnia* both that the availability and acquisition of high-quality food is an important factor in determining ecological interactions and

dynamics (Acharya et al. 2004; Weider et al. 2005), and that swimming at both the small and large scale is energetically costly (Lampert 1989). This is a reasonable result given the centrality of metabolic variation to size and swimming behavior variation. Interestingly, no globins and only a single gene with antioxidant activity are associated with PC1, even though CN1 recovered oxygen transport and antioxidant activity term enrichment. The reality that complex quantitative traits—and likely the common genetic variation among suites of traits—are underlain by variation in a large number of genes and loci (Rockman 2008; Mackay et al. 2009) suggests that the larger number of genes for CN2 should not be unexpected.

In addition to the biological and statistical cohesion of coexpression modules, the recovery of promoter motifs enriched among the coexpression network genes lends additional support to the network. Identifying the target sequences of transcriptional regulators is an important aspect of understanding the evolution of gene expression (López-Maury et al. 2008; Hodgins-Davis & Townsend 2009; Ogasawara & Okubo 2009), and more generally for identifying the "loci of evolution" (i.e., the relative roles of cis- and trans-regulation evolution; (Hoekstra & Coyne 2007; Stern & Orgogozo 2009)). The four enriched promoter motifs found among CN1 genes, and nine recovered for CN2, represent computational hypotheses to be empirically tested (e.g., using CHiP-seq). The motifs of CN2 that are over-represented among coexpression module-specific genes are perhaps the top candidates and are complemented with motifs that possess high sequence similarity with known promoters from *Drosophila*. Motif 6 is most-similar ($p = 7.13e^{-3}$) to the binding target of *Drosophila* kni, a transcription factor associated with muscle, epidermis, and other developmental processes by transcriptional repression (FlyBase; http://flybase.org). Motif 9 shows strongest similarity ($p = 4.48e^{-5}$) to the

binding motif of *Drosophila* gene sqz, a transcription factor with experimental evidence for neural development, oogenesis, and muscle organ development (FlyBase; http://flybase.org). Motif 3 possesses strong similarity ($p = 2.38e^{-6}$) to the binding motif of *Drosophila* gene *Trl*, a transcription factor with experimental evidence for a variety of developmental and metabolic processes (FlyBase; http://flybase.org). Should these promoters be the targets in *Daphnia* of the transcription factors they are targets of in *Drosophila*, a prominent role for developmental regulation genes is suggested for *Daphnia*'s keystone status.

While we have identified two candidate keystone gene networks associated with variation in *Daphnia* traits known to affect the species' ecological requirements and impacts, two primary limitations to our approach must be recognized. First, the expression and phenotype data were collected in common garden conditions to capture constitutive variation, but each of the three D. *pulex* (group) ecotypes would typically experience different "constitutive" conditions in nature. For example, baseline conditions for classic D. *pulex* would entail no fish predators but high-quality (i.e., high phosphate) food, while classic D. *pulicaria* would experience chemical cues from fish predators and a mixture of high- and low-quality food. By focusing on a single, benign environment in which the 32 *Daphnia* clones were assayed, expression and phenotype differences are due to only genotypic differences, which include both direct genetic effects and differential environmental sensitivity. As such, the current research sets a baseline on which future work—e.g., across a variety of environments—can build. For example, combining these results with recent work examining the *Daphnia* transcriptional response to variation in food quality (Dudycha et al. 2012) or response to toxic cyanobacteria (Asselman et al. 2012) will build a more-complete picture of the relationships between

genetic and phenotypic variation, and hence the molecular basis of *Daphnia*'s role in community variation.

A second limitation is that genetic experiments are expected to be challenging in community ecological genomics, and may prove particularly difficult with *Daphnia*: we have not been able to create crosses from lines in this panel and the task has been very challenging for others (M. Pfrender, pers. comm.). That is, the lack of experimental crosses limits our ability to experimentally randomize the genetic background. First, this limits our ability to isolate the additive component of genetic variance, which is the fuel of evolution between years, and thus to determine the relative contributions of other variance components on trait variation. Second, quantitative trait locus mapping requires a mapping population constructed from crosses of two or more genetically divergent lines, and is a more powerful method than association methods such as those used here. Although RNAi has recently been developed for *Daphnia* (Kato et al. 2011), the success rate is currently very low and transfer to natural or semi-natural experimental contexts will introduce new challenges. Other approaches, such as identifying the signature of selection at loci in the genomic regions containing these genes from large field collections (Pfrender 2012), or mutagenesis/mutation accumulation screens, may prove useful in supporting, rejecting, or adding to the set of genes and loci constituting the keystone network.

CONCLUSIONS AND PROSPECTS

Here we have leveraged the well-known ecology of *Daphnia pulex* and its recently-developed genomic resources to identify two competing hypotheses for a keystone gene network—a gene network for which the effects of variation extend to

entire ecological communities. First we showed there is substantial genetic variation shared among the three traits that shape *Daphnia*'s niche, and that this axis of variation should be largely concordant with the common axis of selection to which *Daphnia* is subject. Second, we identified two candidate gene networks, as well as promoter motifs and SNPs, that offer contrasting views of the genes and molecular processes most important to shaping the keystone traits. These various lines of evidence point to two possibilities for a central network that, if perturbed by environmental or evolutionary changes, should have effects beyond the organism to the shape ecological communities.

Several future directions are possible as the field of "ecological genomics" continues to develop. With respect to *Daphnia* in particular, one possible direction is to experimentally assess the candidate keystone networks by assembling communities of clones with variation in the keystone network expression and selecting replicate populations in mesocosm experiments that have been a mainstay of aquatic community ecology. If variation in the keystone network genes (or regulators of the keystone network) is as central to shaping *Daphnia* ecology as the present analysis suggests, then we should be able to predict the clones that perform best in each environment, as well as the implications for evolutionary, population, and community dynamics. A logical complement is employing additional methods to determine or confirm (in the case of the TFs and other genes discussed above) the genetic variants driving keystone network variation.

More generally, we suggest a framework for linking genome biology to ecology by emphasizing the role of keystone species in complex systems. Many communities may be shaped by one or a small number of keystone species whose impacts are disproportionately large given their abundance. Efforts to identify the molecular basis of

variation in ecologically important traits in other keystone species will likely uncover keystone genes and gene networks that explain a substantial amount of variation in community variation, beyond the freshwater systems shaped by *Daphnia* and its keystone network.

METHODS

Details of the methods used to collect the core data used in these analyses is provided in Chapters 1 and 2. We recap those methods before describing the methods specific to this analysis.

Background methods

Daphnia pulex-group clones were collected from a variety of waterbodies in the area surrounding Kellogg Biological Station, Michigan, USA, in April 2009. The panel was created by single-female isolation and subsequent rearing at the University of Texas at Austin (UT) in 400ml ADaM medium (Klüttgen et al. 1994) in a 20°C, 16:8 L:D environmental chamber, using live *Scenedesmus acutus* and Shellfish Diet (Reed Mariculture, Campbell, CA) for food.

Phenotypes were assayed in three groups of common-garden experiments. Growth and fecundity traits were quantified using three replicates of each clone raised individually through three generations of single-female rearing to remove maternal effects. During the third generation, offspring were counted in each of the first three broods, and collected from the first and third broods for subsequent drying and weighing (to the nearest microgram) on a Sartorius SE-2 ultramicrobalance. At the third brood the

mother was collected, external measurements made with a microscale and subsequently dried and weighed.

Phototaxis was quantified in an assay derived from that of De Meester & Cousyn (1997). Three replicates of ten individuals of each clone were placed in a 30-cm tall glass tube filled with fresh ADaM or ADaM from a tank holding fish (Pimephales promelas) so that kairomones were present. Two milliliters of dilute Shellfish Diet was added to each tube as food. Over a four-hour exposure period the number of individuals in the upper 14cm (U), middle 12cm (M), and lower 4cm (L) of the tube was recorded, with the tube lit from above with a 5000oK fluorescent lamp, and the phototaxis index calculated by:

$$PI = (U - L) / (L + M + U)$$

Phototaxis plasticity was calculated by subtracting $PI_{kairomone}$ from $PI_{control}$.

Micro-scale swimming behaviors were assayed in a 10 x 10 x 1cm glass tank filled with ADaM. Three replicates of a single individual of each clone was placed in the tank, allowed to acclimate for ten minutes, then its swimming behavior was recorded with a Sony MODEL HD camcorder for five minutes under benign conditions. At the five minute mark, a predator "attack" was simulated by quickly and directly prodding the individual with a probe; recording continued five additional minutes to capture the swimming response. We used Kinovea (http://www.kinovea.org) software to extract pre- and post-attack mean swimming velocity, max swimming velocity, the coefficient of variation of swimming velocity, the area occupied per unit time (i.e., given the minimum and maximum x and y coordinates), and the sinuosity of swimming behavior (i.e., total path length per unit area). From these we calculated the plasticity of each trait by subtracting the post-attack values from the pre-attack values. All assays were conducted

between 11:00 and 13:00h local time to coincide with the collection of samples for RNA-seq.

Genomic DNA was extracted from a pool of three individuals of each clone and used to prepare libraries for 2b-RAD genotyping per Wang et al. (2012). RNA was extracted from single individuals raised in three replicates at low density (5 clones $120ml^{-1}$) for three generations to remove maternal and grand-maternal effects. Individuals were collected in random order between 11:00 and 13:00h local time on two subsequent days to minimize circadian artifacts, and each handled ~1.5 minutes before submersion in liquid nitrogen to avoid disturbance effects on expression. Libraries for RNA-seq were prepared for three biological replicates of each clone per standard methods (Meyer et al. 2011). Sequencing was done on ABI SOLiD v4 and 5500XL sequencers at the UT Genome Sequencing and Analysis Facility.

Reads were quality filtered, mapped using SHRiMP 2.2.3, and genotypes were called from 2b-RAD and RNA-seq data using the GATK (DePristo et al. 2011). An advantage of RNA-seq over microarray analysis is the ability to determine which alleles are being expressed, however, allele-specific expression is a hazard for genotyping from RNA-seq data. The signature of allele-specific expression in this data would be heterozygous genotype calls in RAD data and homozygous calls (of one of the two het alleles) in the RNA data. Reads from 2b-RAD and RNA data overlapped at an average of 7122 loci per clone, with an average of 10.3 differences between the genotypes called (0.1%), and there was no enrichment for heterozygotes from 2b-RAD-derived genotypes relative to RNA-seq-derived genotype calls. Given the lack of population genomic resources for *Daphnia* at this time and the extremely low discordance between genotype calls from the two data sources, we elected to pool the data for genotype characterization.

Gene expression was quantified using custom scripts that allocated multi-mapped reads (Ji et al. 2011), down-weighted expression in inverse proportion to expression levels (Taub et al 2010), and normalized expression to reads per million mapped. Samples with < $1e^5$ reads were removed from further analysis, but all clones possessed at least two replicates after filtering. Only genes with an average expression ≥ 5 reads per million mapped were retained for further analysis to reduce conflating non-expressed and non-detected genes. We tested if each gene was differentially expressed (DEG) using a generalized linear model (GLM) with quasipoisson errors (McCullagh & Nelder 1999) with the variance inflation factor estimated for each gene from the pooled samples, and Benjamini-Hochberg (Ewens & Grant 2010) false discovery rate (FDR) control exercised at a 1% rate.

Keystone network analyses

Given the core data, we used two methods to address the problem of identifying candidate keystone gene networks: Candidate Network 1 (CN1) was derived by intersecting the gene sets for each of the three niche-mediating traits; and Candidate Network 2 (CN2) was identified by finding genes and SNPs correlated with the common genetic variance of the three niche-mediating traits. Specifically, for CN1 we extracted the set of DEGs significantly associated with size, swimming speed, and phototaxis from the trait-wise analyses of Chapters 1 and 2, then calculated the two- and three-trait intersections of these gene sets. Expression data for the genes in the union of the intersections was analyzed using WGCNA (Langfelder & Horvath 2008) to identify patterns of coexpression (i.e., to identify coexpression modules).

Identifying CN2 required additional analyses because we did not draw directly from existing results. First, we identified the major axes of genetic variation by PCA decomposition of clone means (i.e., genotypic values) for the three focal traits using the `FactoMineR` package of R 2.15.3 (R Development Core Team 2009). We used the *dimdesc* function of FactoMineR to extract the correlations of each PC with variation in each trait. The strength of association between DEGs and the first eigenvector was quantified using distance correlation (Székely et al. 2007; Székely & Rizzo 2013), which is not dependent on linear or monotonic relationships between variables, and calculated the p-value with the *dcor.ttest* function of the `energy` R package. These p-values were Benjamini-Hochberg adjusted using R's built-in *p.adjust* function and genes retained if $p_{adjusted} < 1e^{-3}$.

We used the `WGCNA` R package to infer the coexpression network and modules of DEGs associated with PC1. Keeping with the methods of Chapters 1 and 2, the residuals of expression ~ PC regressions were used for WGCNA clustering to remove likely spurious relationships (Ayroles et al. 2009), and genes that clustered with the "grey" module in trait-wise coexpression analysis were removed from further consideration as likely spurious relationships. The modularity of coexpression networks may be sensitive to parameter choice, and to determine the "best" network we tested three sets of parameters in WGCNA, including difference combinations of minimum module size and module merge height. Once the coexpression modules were defined for each candidate network, we used the GOstats R package to test for Gene Ontology (GO) term enrichment given universes of all *Daphnia* genes with annotations and DEGs with annotations. We chose the candidate network whose parameters resulted in modules with the strongest and greatest number of enriched GO terms, on the grounds that functionally

similar genes should increase enrichment while poor clustering should result in weaker enrichment. We further examined clustering heatmaps to ensure that the (unsupervised) module assignments of the chosen network were reasonable.

For both CN1 and CN2 we used XXmotif (Hartmann et al. 2013) to identify promoter motifs shared among the DEGs of the keystone network at an effective $E < 1e^{-3}$ (see Hartmann et al. 2013 SI 1). Parameters included searching both strands, using a second-order background model, and an intermediate threshold for merging similar motifs. Given the significant motifs we searched for orthologous motifs in *Drosophila*, which have been identified by, e.g., chromatin immunoprecipitation assays, using TOMTOM (Gupta et al. 2007).

FIGURES

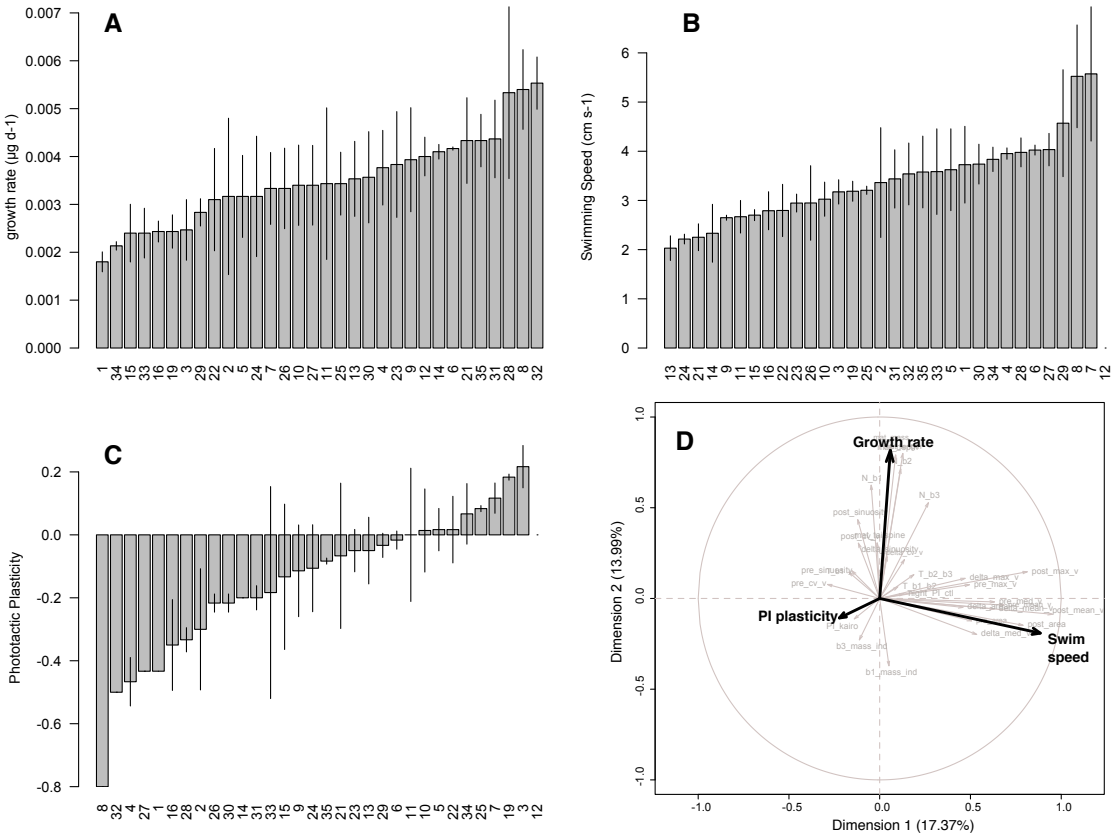


Figure 4.1. Variation in three traits central to *Daphnia* ecology. (A) Growth rate from birth to the release of the third brood; (B) swimming speed (median, post-attack); (C) phototaxis plasticity (constitutive phototaxis index (PI) – kairomone-exposed PI); and (D) the first two principal components of all phenotype data, with the three central traits highlighted. Note that the PC loadings discussed in the text refer to the context of the three keystone traits (SI Figure 4.1). Error bars in A-C are ± 1.96 s.e.m.

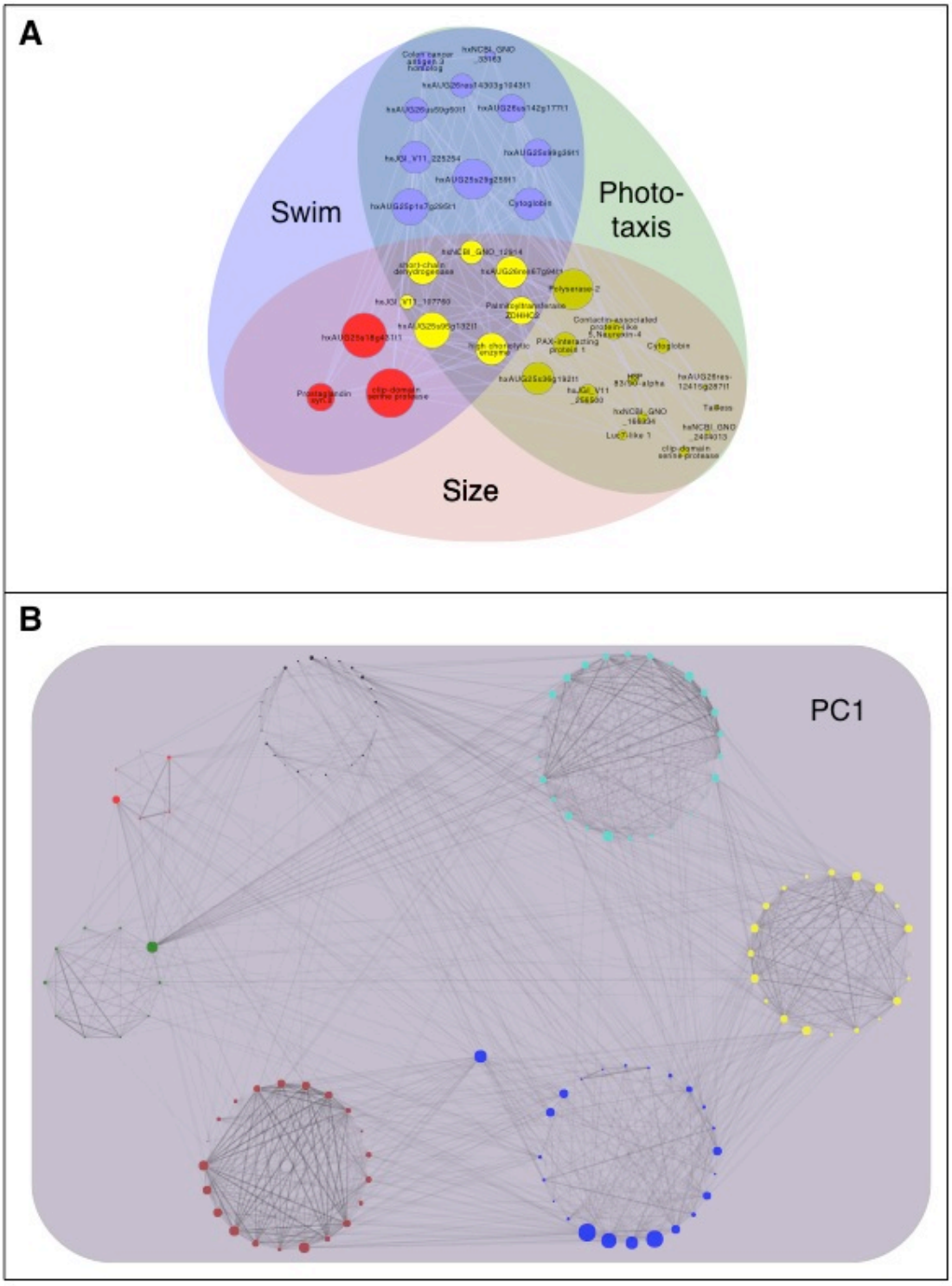

A
Swim
Photo-
taxis
Size
B
PC1

Figure 4.2. The candidate "keystone gene networks" for *Daphnia*. (Previous page) (A) Candidate Network 1, the 34-gene network derived by finding the intersection of genes associated with each niche-mediating trait individually. (B) Candidate Network 2, the 135-gene network of genes whose expression was significantly associated with variation in the first principal component (PC1) of niche-mediating trait variation. In both networks, node size is proportional to the gene's betweenness-centrality (i.e. the relative importance of a node for paths crossing between pairs of nodes), and edge weights are proportional to the absolute value of the expression correlation between connected genes. A handful of high betweenness-centrality genes—those with high connectance between modules—are labeled and discussed in the text.

TABLES

Gene ID	Name	Location	GO terms
hxAUG25p1s7g295t1	YLP motif-containing protein 1/sw	scaffold_7:1423306-1427144:-	GO:0032204/P:regulation of telomere maintenance; GO:0001078/F:RNA pol II proximal region sequence-specific DNA binding TF activity involved in negative regulation of transcription
hxAUG26rep1s4g157t1	zinc finger protein/ARP2_G3163	scaffold_4:1006322-1008720:-	GO:0005488/F:binding,GO:0005575/C:cellular_component,GO:0006464/P:protein modification process,GO:0016740/F:transferase activity
hxAUG26res67g94t1	expressed protein	scaffold_67:449247-449819:+	.

Table 4.1. Differentially expressed genes shared between Candidate Networks 1 and 2. "GO terms" are Gene Ontology terms.

Candidate Network	GOID	P-value	Odds Ratio	Expected Count	Observed Count	Term Size	Term	GO Domain
1	GO:0016209	0.001	21.83	0.17	3	131	antioxidant activity	MF
	GO:0019825	0.002	38.50	0.06	2	47	oxygen binding	MF
	GO:0008233	0.056	3.26	1.51	4	1138	peptidase activity	MF
	GO:0006950	0.005	6.03	1.19	5	1086	response to stress	BP
	GO:0050896	0.025	3.52	2.46	6	2256	response to stimulus	BP
	GO:0019538	0.083	2.48	3.25	6	2980	protein metabolic process	BP
	GO:0005576	0.016	4.24	1.59	5	1242	extracellular region	CC
2	GO:0008233	0.055	2.11	4.22	8	1138	peptidase activity	MF
	GO:0045182	0.061	17.17	0.06	1	17	translation regulator activity	MF
	GO:0016787	0.071	1.71	10.32	15	2783	hydrolase activity	MF
	GO:0006629	0.011	3.47	1.90	6	478	lipid metabolic process	BP
	GO:0044238	0.092	1.51	19.85	25	4990	primary metabolic process	BP
	GO:0005576	0.002	2.99	5.29	13	1242	extracellular region	CC
	GO:0044421	0.003	4.08	1.94	7	456	extracellular region part	CC
	GO:0031012	0.003	7.03	0.63	4	147	extracellular matrix	CC
	GO:0005578	0.003	7.03	0.63	4	147	proteinaceous extracellular matrix	CC

Table 4.2. Gene Ontology term enrichment among the genes of Candidate Networks 1 and 2.

Candidate network	Motif #	Motif consensus	Motif E-val	# key. genes	*Drosophila* matches				
					***Dmel* ID**	***p*-value**	**E-value**	***q*-value**	**Target consensus**
1	Motif_1	GGNSWVSTBGGMGARTA	5.57E-14	1	--	--	--	--	--
	Motif_2	GRGGGKGGKRGGRDRGG	4.83E-03	5	FBgn0086910	9.72E-10	7.89E-07	1.56E-06	GGGGGGGGGGGGGACGTGT
					FBgn0086910_2	6.49E-08	5.26E-05	5.19E-05	GGGGGGGGGGGGGGG
					FBgn0020309_2	1.41E-07	0.000115	7.53E-05	GGGTGGGGGGGGGG
					FBgn0036179_2	1.99E-07	0.000161	7.96E-05	TAGGGGGTGGGGGGG
					FBgn0036179	5.12E-07	0.000415	0.000164	GGGGGGGGGGGGGGGGAAAAT
					FBgn0020309	1.64E-06	0.00133	0.000437	GGGGAGGGGGGGGGT
					FBgn0005630_19	3.47E-05	0.028161	0.007937	GGTGGGGGGCTGGGG
					FBgn0013469_2	6.04E-05	0.048988	0.012081	TGCGTGGGTGGGGGG
					FBgn0003499_2	8.07E-05	0.065435	0.014344	GTGCGTGGGCGGGG
	Motif_3	SGGCSRKRBMGGC	1.93E-02	5	--	--	--	--	--
	Motif_4	MKHDGRSGAYGBYG	2.86E-02	11	--	--	--	--	--
	Motif_5	CGTCGDCDGCK	2.96E-02	8	--	--	--	--	--
2	Motif_1	GGGGGGGGGGGGGGA	3.69E-13	9	FBgn0036179	7.43E-11	6.02E-08	1.18E-07	GGGGGGGGGGGGGGGAAAAT
					FBgn0086910	1.26E-09	1.02E-06	1.00E-06	GGGGGGGGGGGGGACGTGT
					FBgn0036179_2	3.03E-09	2.46E-06	1.61E-06	TAGGGGGTGGGGGGG
					FBgn0020309_2	4.83E-09	3.91E-06	1.92E-06	GGGTGGGGGGGGGGG
					FBgn0020309	8.17E-09	6.62E-06	2.60E-06	GGGGAGGGGGGGGGT
					FBgn0086910_2	1.68E-08	1.36E-05	4.46E-06	GGGGGGGGGGGGGGG
					FBgn0003002_3	1.43E-05	0.012	0.003	CCGCGGGGGGTCTGG
					FBgn0005630_19	5.84E-05	0.047	0.012	GGTGGGGGGCTGGGG
					FBgn0003499	7.90E-05	0.064	0.014	GTGCGTGGGCGGGG
					FBgn0003002_2	9.44E-05	0.077	0.015	CGGTGGGGGGGAT
	Motif_2	AAAAGAAAAAAAAAA	6.62E-12	38	FBgn0027339_2	4.12E-05	0.033	0.067	ACAAAAAAACAAAAA
	Motif_3	AGAGAGAGAGAGAA	5.55E-07	9	Trl	1.94E-06	0.002	0.003	AGAGAGAGAGAAA
					FBgn0013263	4.78E-05	0.039	0.026	GAGAGAGCAA
	Motif 4	SCAGCAGCAR	3.96E-06	26	--	--	--	--	--
	Motif_5	CGCCATCTAGCGGT	6.03E-06	11	FBgn0035997	5.40E-05	0.044	0.087	CGCCATCTTG
	Motif 6	AGAAGARGRRGR	1.33E-05	15	--	--	--	--	--
	Motif_7	TGTGTGTGTGTGTA	8.00E-05	14	FBgn0013469	9.34E-05	0.076	0.144	TGCGTGGGTGG
	Motif_8	TTTTTTTTATTT	3.95E-04	65	FBgn0039350	1.40E-06	0.001	0.002	TTTTTTTTATT
					FBgn0010768	6.53E-05	0.053	0.052	TTTTTTTTTATGTAGC
	Motif_9	AAAAAAAGAAAAAAAAA	7.12E-04	15	FBgn0027339_2	6.55E-05	0.053	0.078	ACAAAAAAACAAAAA
					FBgn0010768	9.69E-05	0.079	0.078	GCTACATAAAAAAAAA

Table 4.3. Promoter motifs enriched among the genes of Candidate Networks 1 and 2. Highly similar known promoter motifs from *Drosophila melanogaster* are provided for comparison.

LITERATURE CITED

Acharya, K., Kyle, M. & Elser, J.J. (2004). Effects of stoichiometric dietary mixing on *Daphnia* growth and reproduction. *Oecologia*, 138, 333–340.

Arnold, S.J., Burger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. (2008). Understanding The Evolution And Stability Of The G-Matrix. *Evolution*, 62, 2451–2461.

Asselman, J., De Coninck, D.I.M., Glaholt, S., Colbourne, J.K., Janssen, C.R., Shaw, J.R., *et al.* (2012). Identification of Pathways, Gene Networks, and Paralogous Gene Families in *Daphnia pulex* Responding to Exposure to the Toxic Cyanobacterium Microcystis aeruginosa. *Environ. Sci. Technol.*, 46, 8448–8457.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., *et al.* (2009). Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet*, 41, 299–307.

Biesecker, L.G. (2013). Hypothesis-generating research and predictive medicine. *Genome Res.*, 23, 1051–1053.

Boeing, W.J., Ramcharan, C.W. & Riessen, H.P. (2006). Multiple predator defence strategies in *Daphnia pulex* and their relation to native habitat. *J. Plankton Res.*, 28, 571 –584.

Brooks, J.L. & Dodson, S.I. (1965). Predation, Body Size, and Composition of Plankton. *Science*, 150, 28.

Carpenter, S.R. & Kitchell, J.F. (1996). *The trophic cascade in lakes*. Cambridge University Press.

Carroll, S., Hendry, A., Reznick, D. & Fox, C. (2007). Evolution on ecological time-scales. *Funct. Ecol.*, 21, 387–393.

Cernỳ, M. & Bytel, J. (1991). Density and size distribution of *Daphnia* populations at different fish predation levels. *Hydrobiologia*, 225, 199–208.

Chamberlin, T.C. (1890). The method of multiple working hypotheses. *Science*, 15, 92–96.

Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University Of Chicago Press.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., *et al.* (2011). The Ecoresponsive Genome of *Daphnia pulex*. *Science*, 331, 555 –561.

Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford University Press, USA.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, 43, 491–498.

Dini, M.L. & Carpenter, S.R. (1992). Fish predators, food availability and diel vertical migration in *Daphnia*. *J. Plankton Res.*, 14, 359–377.

Dodson, S. & Ramcharan, C. (1991). Size-specific swimming behavior of *Daphnia pulex*. *J. Plankton Res.*, 13, 1367 –1379.

Dudycha, J.L., Brandon, C.S. & Deitz, K.C. (2012). Population genomics of resource exploitation: insights from gene expression profiles of two *Daphnia* ecotypes fed alternate resources. *Ecol. Evol.*

Ellner, S.P., Geber, M.A. & Hairston Jr, N.G. (2011). Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. *Ecol. Lett.*

Ewens, W.J. & Grant, G.R. (2010). *Statistical Methods in Bioinformatics: An Introduction*. Springer.

Feder, M.E. & Mitchell-Olds, T. (2003). Evolutionary and ecological functional genomics. *Nat. Rev. Genet.*, 4, 649–655.

Forbes, S.A. (1880). The Food of the Darters. *Am. Nat.*, 14, 697–703.

Gliwicz, Z. & Umana, G. (1994). Cladoceran body-size and vulnerability to copepod predation. *Limnol. Ocean.*, 39, 419–424.

Goulden, C.E., Henry, L.L. & Tessier, A.J. (1982). Body Size, Energy Reserves, and Competitive Ability in Three Species of Cladocera. *Ecology*, 63, 1780–1789.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biol.*, 8, R24.

Hairston, N., Lampert, W., Caceres, C., Holtmeier, C., Weider, L., Gaedke, U., *et al.* (1999). Lake ecosystems - Rapid evolution revealed by dormant eggs. *NATURE*, 401, 446–446.

Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T. & Fox, J.A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.*, 8, 1114–1127.

Hartmann, H., Guthöhrlein, E.W., Siebert, M., Luehr, S. & Söding, J. (2013). P-value-based regulatory motif discovery using positional weight matrices. *Genome Res.*, 23, 181–194.

Hodgins-Davis, A. & Townsend, J.P. (2009). Evolving gene expression: from G to E to G× E. *Trends Ecol. Evol.*, 24, 649–658.

Hoekstra, H.E. & Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, 61, 995–1016.

Hood, J.M. & Sterner, R.W. (2010). Diet Mixing: Do Animals Integrate Growth or Resources across Temporal Heterogeneity? *Am. Nat.*, 176, 651–663.

Ji, Y., Xu, Y., Zhang, Q., Tsui, K.-W., Yuan, Y., Norris Jr., C., *et al.* (2011). BM-Map: Bayesian Mapping of Multireads for Next-Generation Sequencing Data. *Biometrics*, 67, 1215–1224.

Kato, Y., Shiga, Y., Kobayashi, K., Tokishita, S., Yamagata, H., Iguchi, T., *et al.* (2011). Development of an RNA interference method in the cladoceran crustacean *Daphnia* magna. *Dev. Genes Evol.*

Kerfoot, W.C., Kellogg Jr, D.L. & Strickler, J.R. (1980). Visual observations of live zooplankters: evasion, escape, and chemical defenses. *Evol. Ecol. Zooplankton Communities Univ. Press New Engl.*, 10–27.

Kirkpatrick, M. (2009). Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, 136, 271–284.

Klüttgen, B., Dülmer, U., Engels, M. & Ratte, H.. (1994). ADaM, an artificial freshwater for the culture of zooplankton. *Water Res.*, 28, 743–746.

Lampert, W. (1989). The Adaptive Significance of Diel Vertical Migration of Zooplankton. *Funct. Ecol.*, 3, 21–27.

Lampert, W. (2006). *Daphnia*: model herbivore, predator and prey. *Pol. J. Ecol.*, 54, 607–620.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, 402–416.

Lande, R. & Arnold, S.J. (1983). The measurement of selection on correlated characters. *Evolution*, 1210–1226.

Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.

Leibold, M. & Tessier, A.J. (1991). Contrasting patterns of body size for *Daphnia* species that segregate by habitat. *Oecologia*, 86, 342–348.

Leibold, M.A. (1989). Resource edibility and the effects of predators and productivity on the outcome of trophic interactions. *Am Nat*, 134, 922.

Leibold, M.A. (1996). A graphical model of keystone predators in food webs: Trophic regulation of abundance, incidence, and diversity patterns in communities. *Am. Nat.*, 147, 784–812.

López-Maury, L., Marguerat, S. & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.*, 9, 583–593.

Lovell, J.T., Juenger, T.E., Michaels, S.D., Lasky, J.R., Platt, A., Richards, J.H., *et al.* (2013). Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proc. R. Soc. B Biol. Sci.*, 280.

Mackay, T.F.C. (2001). The genetic architecture of quantitative traits. *Annu. Rev. Genet.*, 35, 303–339.

Mackay, T.F.C., Stone, E.A. & Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet*, 10, 565–577.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Candidate loci, genes, and networks underlying ecologically important behavioral variation in a model ecological species, *Daphnia pulex*.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Exploring the molecular basis of variation in size and fecundity in an emerging model organism, *Daphnia pulex*.

McCullagh, P. & Nelder, J.A. (1999). *Generalized linear models*. Chapman & Hall, CRC.

De Meester, L. (1991). An analysis of the phototactic behaviour of *Daphnia* magna clones and their sexual descendants. *Hydrobiologia*, 225, 217–227.

De Meester, L. & Cousyn, C. (1997). The change in phototactic behaviour of a *Daphnia* magna clone in the presence of fish kairomones: the effect of exposure time. *Hydrobiologia*, 360, 169–175.

De Meester, L., Gómez, A., Okamura, B. & Schwenk, K. (2002). The Monopolization Hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta Oecologica*, 23, 121–135.

De Meester, L. & Weider, L.J. (1999). Depth selection behavior, fish kairomones, and the life histories of *Daphnia* hyalina$\times$ galeata hybrid clones. *Limnol. Ocean.*, 1248–1258.

Meyer, E., Aglyamova, G.V. & Matz, M.V. (2011). Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol. Ecol.*, 20, 3599–3616.

Mills, L.S., Soulé, M.E. & Doak, D.F. (1993). The keystone-species concept in ecology and conservation. *BioScience*, 43, 219–224.

Miner, B.E., Meester, L.D., Pfrender, M.E., Lampert, W. & Hairston, N.G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proc. R. Soc. B Biol. Sci.*, 279, 1873–1882.

Odling-Smee, F.J., Laland, K.N. & Feldman, M.W. (2003). *Niche Construction: The Neglected Process in Evolution (MPB-37)*. Princeton University Press.

Ogasawara, O. & Okubo, K. (2009). On Theoretical Models of Gene Expression Evolution with Random Genetic Drift and Natural Selection. *PLoS ONE*, 4, e7943.

Paine, R.T. (1969). A note on trophic complexity and community stability. *Am. Nat.*, 103, 91–93.

Persson, J., Brett, M.T., Vrede, T. & Ravet, J.L. (2007). Food quantity and quality regulation of trophic transfer between primary producers and a keystone grazer (*Daphnia*) in pelagic freshwater food webs. *Oikos*, 116, 1152–1163.

Pfrender, M.E. (2012). Triangulating the genetic basis of adaptation to multifarious selection. *Mol. Ecol.*, 21, 2051–2053.

Pijanowska, J., Weider, L.J. & Lampert, W. (1993). Predator-mediated genotypic shifts in a prey population: experimental evidence. *Oecologia*, 96, 40–42.

R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rockman, M.V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature*, 456, 738–744.

Sarnelle, O. (1993). Herbivore effects on phytoplankton succession in a eutrophic lake. *Ecol. Monogr.*, 129–149.

Sarnelle, O. (2005). *Daphnia* as keystone predators: effects on phytoplankton diversity and grazing resistance. *J. Plankton Res.*, 27, 1229.

Scoville, A.G. & Pfrender, M.E. (2010). Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. *Proc. Natl. Acad. Sci.*, 107, 4260 –4263.

Seuront, L., Brewer, M.C., Strickler, J.R. & others. (2004). Quantifying zooplankton swimming behavior: the question of scale. *Handb. Scaling Methods Aquat. Ecol. Anal. Simul. CRC Press Boca Raton*, 333–359.

Simon, J.-C., Pfrender, M.E., Tollrian, R., Tagu, D. & Colbourne, J.K. (2011). Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. *J. Hered.*, 102, 512 –525.

Spitze, K. (1991). Chaoborus Predation and Life-History Evolution in *Daphnia pulex*: Temporal Pattern of Population Diversity, Fitness, and Mean Life History. *Evolution*, 45, 82–92.

Stern, D. & Orgogozo, V. (2009). Is genetic evolution predictable? *SCIENCE*, 323, 746.

Székely, G.J. & Rizzo, M.L. (2013). The distance correlation -test of independence in high dimension. *J. Multivar. Anal.*, 117, 193–213.

Székely, G.J., Rizzo, M.L. & Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Stat.*, 35, 2769–2794.

Tessier, A., Young, A. & Leibold, M. (1992). Population dynamics and body-size selection in *Daphnia*. *Limnol. Ocean.*, 37, 1–13.

Wang, S., Meyer, E., McKay, J.K. & Matz, M.V. (2012). 2b-rad: a simple and flexible method for genome-wide genotyping. *Nat. Methods*, 9, 808–810.

Weider, L.J., Makino, W., Acharya, K., Glenn, K.L., Kyle, M., Urabe, J., *et al.* (2005). Genotype$\times$ environment interactions, stoichiometric food quality effects, and clonal coexistence in *Daphnia pulex*. *Oecologia*, 143, 537–547.

Wetzel, R.G. (2001). *Limnology: lake and river ecosystems*. Academic press.

Wolinska, J., Löffler, A. & Spaak, P. (2007). Taxon‐specific reaction norms to predator cues in a hybrid *Daphnia* complex. *Freshw. Biol.*, 52, 1198–1209.

Conclusion: Ecological Genomics Of *Daphnia pulex*

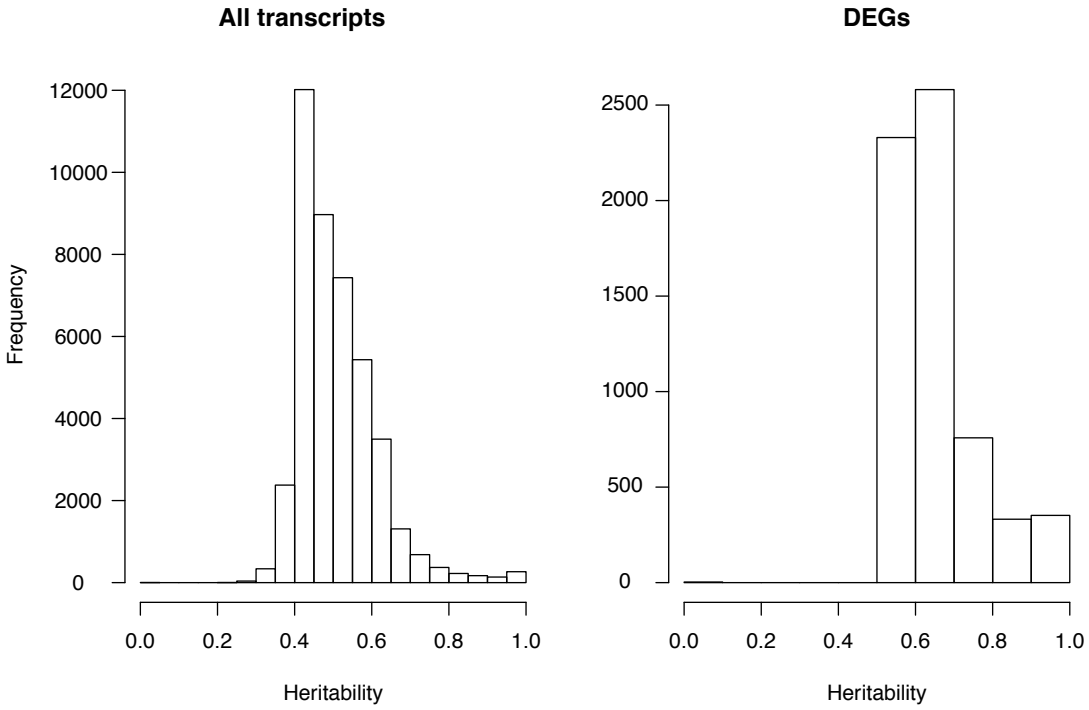
Determining the molecular basis of variation in complex quantitative traits is a challenging problem, and to succeed at this task is (or will be) a substantial accomplishment. Setting a goal of explaining variation across several levels of organization, from genotype to community, presents numerous additional challenges in largely uncharted territory. However, the prospect of understanding the interactions between these many levels—to be able to predict states and dynamics at higher levels from lower-level variation, and vice-versa—provides the impetus to take up the challenge. The breadth and depth of this problem extends far beyond a single dissertation, but here I have attempted to contribute to our understanding by focusing on the genetics and genomics of *Daphnia pulex*, a model ecological species that is known to impact the constituency and dynamics of freshwater ponds and lakes. I recovered a large number of genes associated with variation in 32 organismal traits; developed a model for the molecular basis of variation in the offspring size-number trade-off; and proposed two competing hypotheses for "keystone gene networks", networks that may be central to shaping communities because of their role in shaping *Daphnia* ecology. The implications of these results range from the development of hundreds to testable gene-trait hypotheses for *Daphnia*, to a model that suggests different genes and biological pathways underlie variation in the offspring size-number trade-off, to a pair of candidate gene networks that may explain how a 1-3mm long zooplankter can shape entire freshwater pond and lake communities.

Appendix: Supplemental Information

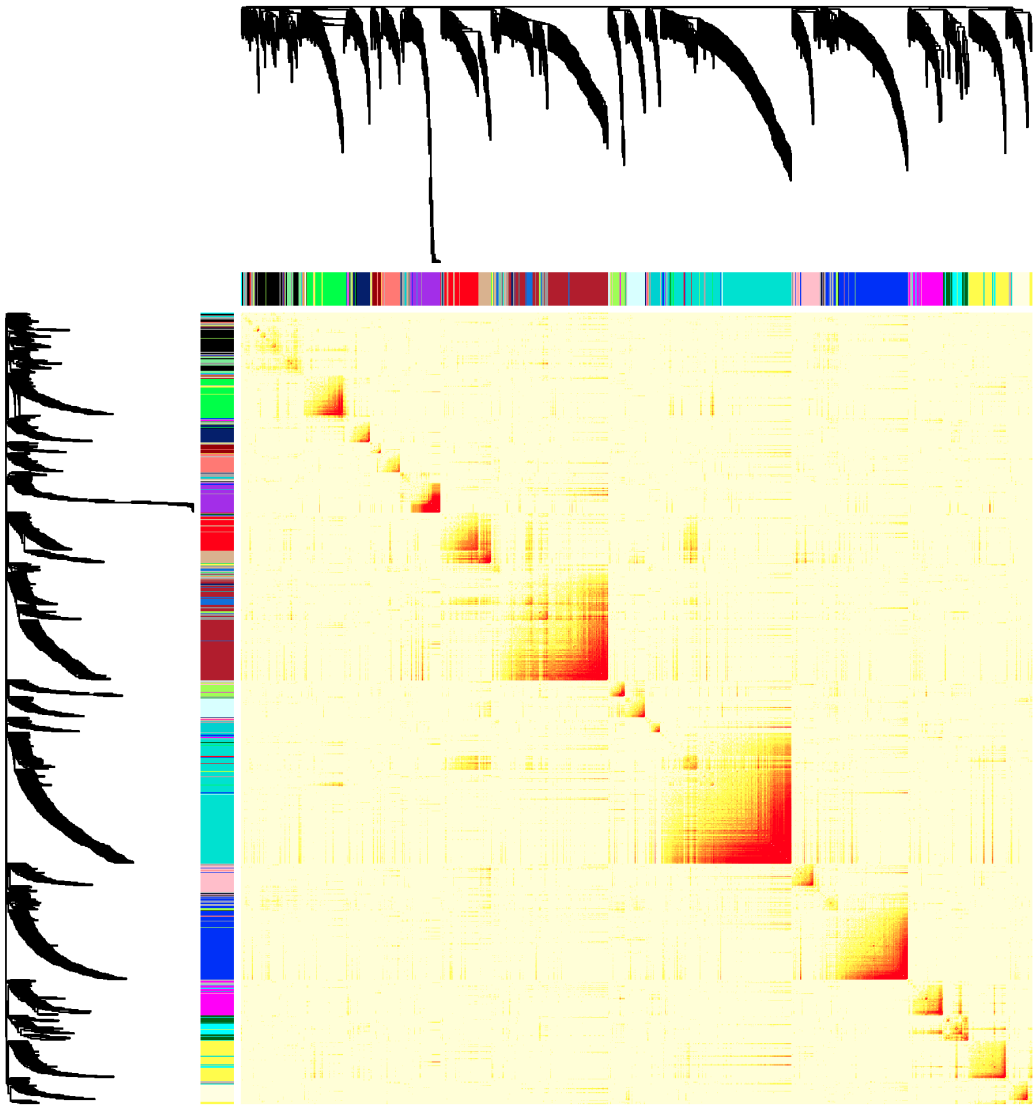
Genomics research provides voluminous results that are often more expansive than can be reasonably printed. A subset of the Supplemental Information (SI) referenced in this dissertation are provided directly in this section; the larger SI Tables are included as digital uploads.

CHAPTER 1 SUPPLEMENTAL INFORMATION

SI Figures



SI Figure 1.1. Distributions of *Daphnia* transcript heritabilities for all transcripts and differentially expressed genes (DEGs).



SI Figure 1.2. Heatmap of global coexpression modules; module colors on the margins are referenced in the main text.

Global module	N_b1	N_b2	N_b3	T_b1	T_b1_b2	T_b2_b3	b1_mass_ind	b3_mass_ind	Growth_rate	mat_depth	mat_length	mat_mass	mat_tailspine
blue	-	11	14	12	17	11	-	19	-	26	-	14	15
green	-	16	-	-	x	-	30	53	20	37	84	-	-
black	-	-	44	-	-	-	-	-	-	-	-	14	-
salmon	-	-	-	-	x	-	-	-	x	-	-	-	29
brown	17	26	13	-	21	27	-	26	x	-	-	52	-
purple	-	-	31	x	-	-	-	-	28	x	-	13	-
turquoise	14	12	-	20	-	x	20	30	-	-	12	-	18
cyan	-	x	-	19	-	x	-	-	-	-	-	-	-
magenta	-	-	-	-	-	-	-	-	x	x	x	x	-
midnightblue	-	12	-	-	-	-	x	-	x	39	42	-	-
grey60	20	25	13	20	-	x	-	17	x	-	11	21	-
darkgrey	-	-	-	x	-	-	-	-	-	x	x	7	9
yellow	-	-	-	-	-	13	-	-	-	-	x	-	-
royalblue	-	-	-	x	x	-	-	-	x	x	x	-	-
tan	-	-	x	-	22	-	x	-	-	-	-	-	-
lightyellow	13	23	-	-	-	x	x	-	x	-	x	-	-
orange	x	-	x	-	x	-	x	x	x	x	x	-	-
greenyellow	-	-	-	-	x	-	11	-	x	-	-	-	-
pink	-	-	x	-	-	-	-	22	-	-	-	-	-
red	-	-	-	-	-	-	-	-	x	-	-	-	24
lightgreen	11	x	-	x	-	-	-	-	x	x	x	-	-
white	-	6	x	11	-	x	x	-	-	-	-	-	-
darkturquoise	-	-	-	-	-	-	-	-	x	-	-	-	-
lightcyan	-	-	x	x	x	-	-	-	x	-	x	x	-
darkgreen	x	x	-	x	x	x	10	x	x	-	-	-	-
darkorange	-	-	x	x	x	x	-	-	x	-	x	x	x
darkred	x	-	-	x	x	x	x	-	-	-	x	x	-
grey	-	-	x	-	-	x	-	-	-	-	-	-	-
# global modules	5	8	5	5	3	3	4	6	2	3	4	6	5
# trait modules	4	3	5	6	5	5	3	3	4	4	5	5	5

SI Figure 1.3. Mappings between trait-associated DEGs and global coexpression modules.

SI Figure 1.4. REVIGO plots of enriched Gene Ontology terms for *Daphnia* life history traits. (Digital upload)

SI Tables

Clone	N_b1	b1_mass_ind	N_b2	N_b3	b3_mass_ind	mat_mass	mat_length	mat_depth	mat_tailspine	T_b1	T_b1_b2	T_b2_b3	Growth_rate
1	6.000	0.005	7.000	6.000	0.005	0.034	1.700	1.150	0.775	8.667	4.667	3.333	0.002
	5.2915	0.0031	4.5826	2.6458	0.0006	0.0020	0.1000	0.0500	0.1250	2.5166	3.2146	1.5275	0.0004
2	2.333	0.004	4.000	4.000	0.002	0.049	1.400	0.900	0.567	9.000	3.333	2.667	0.003
	0.5774	0.0025	3.0000	1.0000	0.0000	0.0411	0.1000	0.1000	0.2082	2.0000	0.5774	0.5774	0.0028
3	10.000	0.008	7.667	15.667	0.002	0.056	2.267	1.450	0.583	14.667	3.333	2.667	0.002
	3.4641	0.0061	8.0829	8.7369	0.0006	0.0178	0.2887	0.0500	0.0289	6.3509	0.5774	0.5774	0.0011
4	7.667	0.006	6.333	13.000	0.003	0.065	2.267	1.567	NA	11.333	3.000	2.000	0.004
	1.5275	0.0053	2.8868	5.5678	0.0006	0.0180	0.2517	0.0577	NA	0.5774	0.0000	0.0000	0.0014
5	3.667	0.006	10.000	15.333	0.004	0.065	2.367	1.500	0.533	13.333	2.667	3.000	0.003
	2.0817	0.0021	3.4641	2.0817	0.0058	0.0239	0.1155	0.2646	0.0577	4.0415	0.5774	0.0000	0.0015
6	9.667	0.002	12.000	20.667	0.002	0.073	2.667	1.700	0.600	12.000	2.333	2.667	0.004
	5.6862	0.0006	2.6458	4.0415	0.0000	0.0090	0.1155	0.1000	0.2000	2.0000	0.5774	0.5774	0.0001
7	5.667	0.004	13.667	11.000	0.002	0.078	2.433	1.467	0.700	12.000	3.667	7.333	0.003
	2.3094	0.0015	6.6583	5.2915	0.0006	0.0247	0.2517	0.2517	0.1000	1.0000	0.5774	4.0415	0.0013
8	3.333	0.003	16.667	24.667	0.003	0.098	2.500	1.567	NA	9.000	5.000	3.000	0.005
	2.3094	0.0015	8.0829	13.3167	0.0010	0.0396	0.4583	0.3215	NA	1.0000	3.4641	1.7321	0.0014
9	9.000	0.003	12.000	16.333	0.004	0.096	2.433	1.700	0.900	18.000	2.667	5.333	0.004
	7.2111	0.0012	9.5394	8.3865	0.0015	0.0145	0.2082	0.1732	0.2000	7.0000	2.0817	2.5166	0.0019
10	11.000	0.007	9.667	14.000	0.004	0.078	2.233	1.433	0.550	14.000	4.000	4.667	0.003
	2.6458	0.0038	6.6583	5.2915	0.0012	0.0171	0.1528	0.2082	0.0500	1.0000	2.6458	1.5275	0.0015
11	11.000	0.005	6.667	10.667	0.002	0.062	2.167	1.567	NA	12.500	1.500	2.500	0.003
	6.0000	0.0030	4.5092	8.3267	0.0000	0.0526	0.5508	0.3215	NA	2.5000	0.5000	0.5000	0.0027
12	13.500	0.004	16.000	22.000	0.004	0.077	2.500	1.550	NA	12.000	3.000	3.500	0.004
	2.5000	0.0015	4.0000	17.0000	0.0015	0.0085	0.2000	0.0500	NA	1.0000	0.0000	2.5000	0.0007
13	13.000	0.003	12.667	13.333	0.002	0.073	2.500	1.533	0.767	10.333	6.667	5.333	0.004
	4.0000	0.0006	4.9329	7.0238	0.0006	0.0150	0.3606	0.1528	0.0577	1.5275	5.0332	4.9329	0.0014
14	6.667	0.003	9.667	10.333	0.002	0.069	2.300	1.400	0.400	11.000	2.667	2.667	0.004
	1.1547	0.0021	6.6583	5.6862	0.0010	0.0140	0.3464	0.1732	0.1000	3.6056	0.5774	0.5774	0.0003
15	4.333	0.004	6.667	10.000	0.003	0.046	2.133	1.267	0.633	11.333	2.667	4.333	0.002
	1.5275	0.0021	5.6862	4.5826	0.0010	0.0157	0.2517	0.1528	0.1155	1.5275	0.5774	2.3094	0.0010
16	7.333	0.001	7.667	6.333	0.001	0.058	2.333	1.433	0.650	14.667	5.667	3.333	0.002
	4.1633	0.0006	5.7735	4.9329	0.0006	0.0084	0.1528	0.0577	0.0500	2.8868	3.0551	0.5774	0.0004
19	5.667	0.003	3.000	6.000	0.004	0.062	2.100	1.333	0.433	18.667	2.667	3.667	0.002
	1.5275	0.0020	2.6458	0.0000	0.0030	0.0140	0.1000	0.1528	0.0577	8.6217	0.5774	1.1547	0.0006
21	3.333	0.003	13.667	10.667	0.002	0.066	2.200	1.500	0.600	9.667	2.000	3.000	0.004
	0.5774	0.0025	8.0829	4.0415	0.0006	0.0195	0.2000	0.2000	0.1000	0.5774	0.0000	0.0000	0.0016
22	11.000	0.001	14.000	9.000	0.002	0.059	2.533	1.700	0.700	12.333	3.000	2.333	0.003
	9.5394	0.0006	5.2915	6.0828	0.0006	0.0448	0.4933	0.3606	0.1000	4.0415	0.0000	0.5774	0.0019
23	10.667	0.002	18.333	5.333	0.004	0.108	2.500	1.533	0.600	15.000	4.000	5.667	0.004
	9.6090	0.0012	16.6533	1.5275	0.0029	0.0770	0.3606	0.3786	0.0000	4.5826	1.7321	4.6188	0.0019
24	6.333	0.003	11.000	11.000	0.004	0.071	2.533	1.467	0.767	12.000	2.667	4.333	0.003
	3.0551	0.0012	4.5826	5.5678	0.0046	0.0611	0.6429	0.4509	0.2309	2.0000	0.5774	3.2146	0.0022
25	9.333	0.002	12.667	11.333	0.004	0.054	2.133	1.433	0.550	9.000	3.000	3.000	0.003
	1.1547	0.0006	5.6862	3.7859	0.0031	0.0145	0.3055	0.1155	0.0866	0.0000	1.0000	1.0000	0.0011
26	8.333	0.002	13.000	13.000	0.002	0.049	2.167	1.300	0.650	9.333	2.000	2.333	0.003
	2.0817	0.0006	13.1149	5.0000	0.0000	0.0235	0.2517	0.2000	0.0500	1.1547	1.0000	0.5774	0.0015
27	9.667	0.003	13.333	13.333	0.005	0.062	2.300	1.433	0.633	11.000	2.667	3.000	0.003
	6.6583	0.0006	14.4684	9.2916	0.0026	0.0314	0.3464	0.2309	0.0577	1.7321	0.5774	1.0000	0.0014
28	10.667	0.003	13.667	12.000	0.003	0.085	2.400	1.567	0.550	11.333	2.667	2.667	0.005
	5.0332	0.0012	11.5036	3.4641	0.0006	0.0355	0.2646	0.3055	0.1500	4.9329	0.5774	0.5774	0.0031
29	7.333	0.002	9.000	14.000	0.002	0.045	2.367	1.367	0.633	9.667	2.667	3.000	0.003
	2.5166	0.0012	3.6056	5.2915	0.0010	0.0076	0.4041	0.1528	0.0577	1.1547	1.5275	0.0000	0.0005
30	4.000	0.003	5.000	18.000	0.003	0.078	2.650	1.550	0.700	12.500	5.000	6.000	0.004
	0.0000	0.0005	0.0000	9.0000	0.0005	0.0230	0.2500	0.1500	0.0000	3.5000	2.0000	3.0000	0.0017
31	15.000	0.003	28.333	21.667	0.003	0.096	2.567	1.433	0.500	15.000	3.667	2.667	0.004
	8.8882	0.0021	13.5769	10.7858	0.0023	0.0327	0.3055	0.0577	0.1000	2.6458	1.1547	0.5774	0.0014
32	14.667	0.002	18.000	20.333	0.002	0.099	2.733	1.700	0.633	11.000	2.667	3.667	0.006
	9.2916	0.0012	13.5277	2.3094	0.0006	0.0246	0.2517	0.2000	0.0577	3.0000	1.5275	2.3094	0.0009
33	7.333	0.003	6.667	NA	0.002	0.047	2.267	1.300	0.800	13.333	2.667	3.000	0.002
	3.0551	0.0006	4.0415	NA	0.0006	0.0167	0.4619	0.1732	0.1000	2.0817	0.5774	1.0000	0.0009
34	4.000	0.005	6.000	6.333	0.005	0.050	2.300	1.367	0.467	14.000	3.333	3.667	0.002
	1.7321	0.0029	6.0828	2.3094	0.0036	0.0080	0.4359	0.2887	0.0577	2.6458	1.5275	1.1547	0.0002
35	7.333	0.003	13.000	11.000	0.001	0.083	2.367	1.533	0.650	11.000	2.667	5.000	0.004
	0.5774	0.0012	3.0000	3.6056	0.0006	0.0222	0.1528	0.1155	0.0500	1.0000	0.5774	2.0000	0.0010

SI Table 1.1. Clone-wise summary statistics of *Daphnia* life history traits; mean value above and standard deviation below.

SI Table 1.2. Promoter motif enrichment for *Daphnia* global coexpression modules. (Digital upload)

SI Table 1.3. Gene Ontology term enrichment for *Daphnia* global coexpression modules. (Digital upload)

SI Table 1.4 Differentially expressed genes associated with *Daphnia* life history trait variation. (Digital upload)

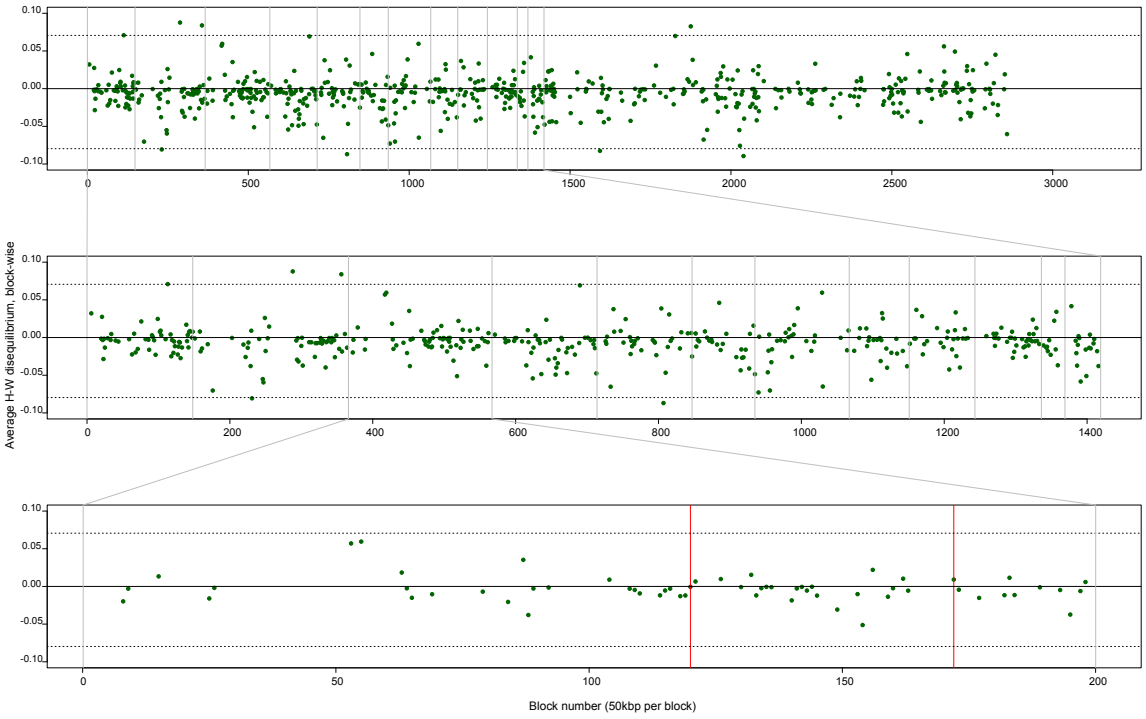
SI Table 1.5 Gene Ontology term enrichment for genes associated with *Daphnia* life history trait variation. (Digital upload)

SI Table 1.6 Promoter motif enrichment for *Daphnia* genes associated with life history trait variation. (Digital upload)

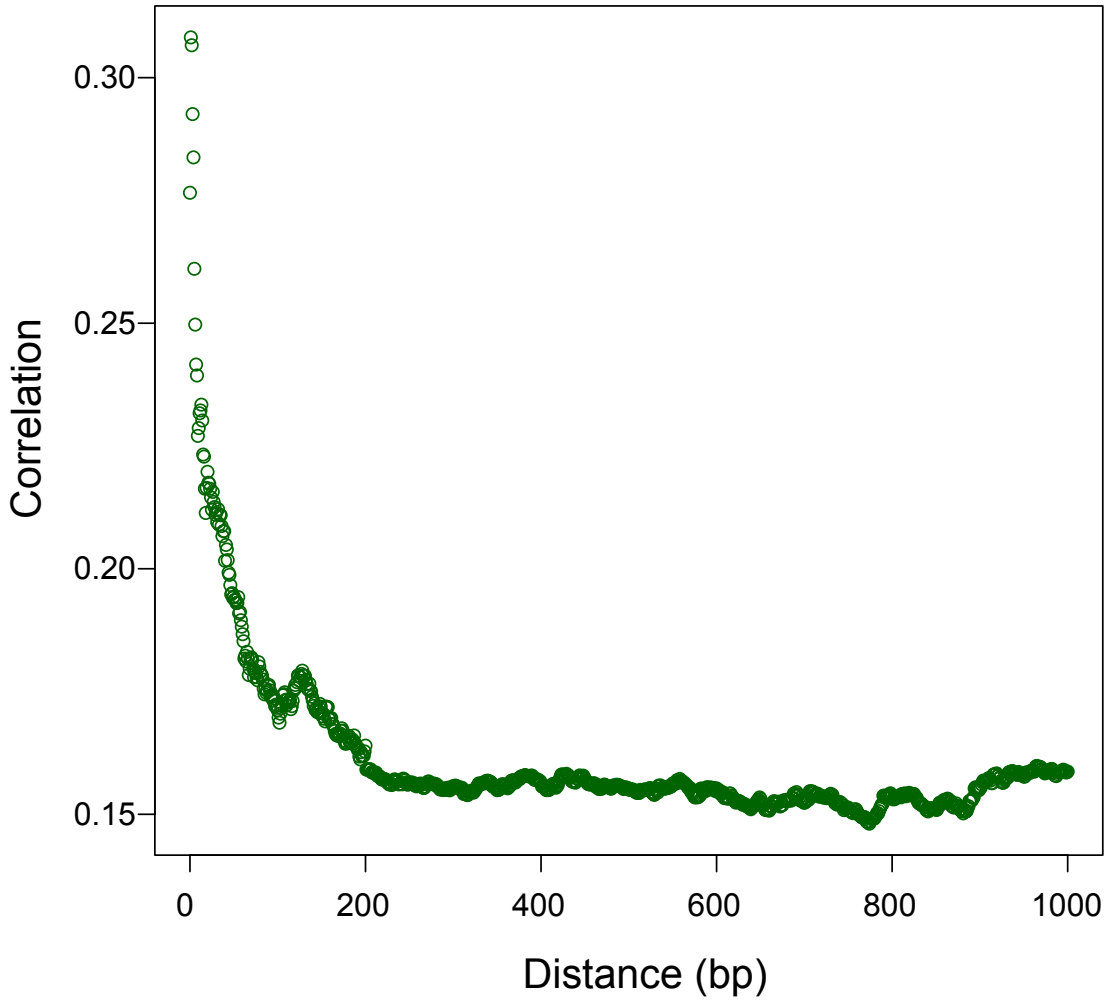
SI Table 1.7 SNPs and differentially expressed genes of an integrated network associated with *Daphnia* life history trait variation. (Digital upload)

CHAPTER 2 SUPPLEMENTAL INFORMATION

SI Figures

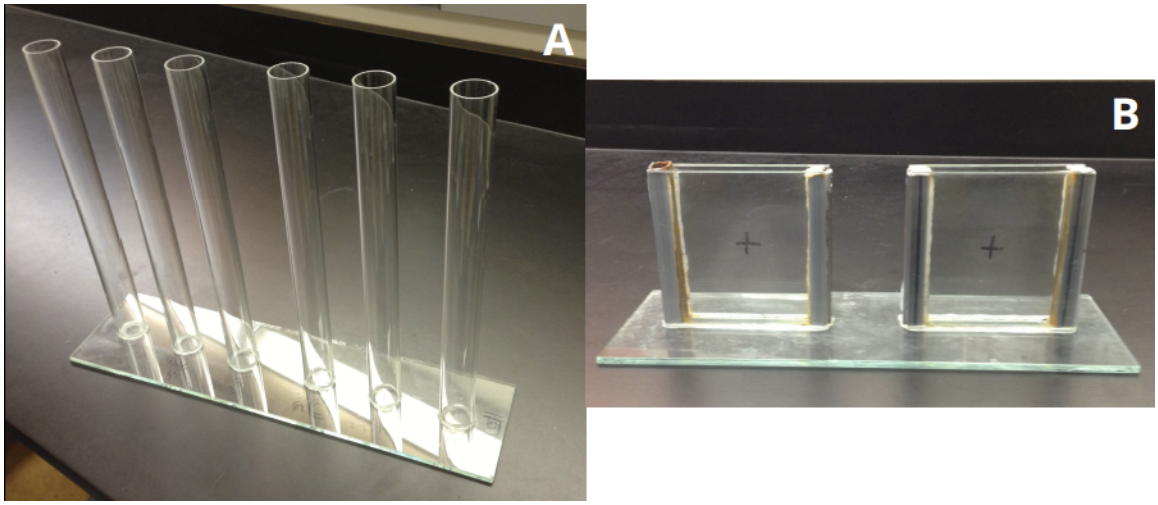


SI Figure 2.1. ***Daphnia*** **genome-wide Hardy-Wienburg disequilibrium (i.e., departure from HWE) in 50kb windows.**



SI Figure 2.2. Linkage decay in the *Daphnia* panel. Values are calculated as a moving average in order to have enough data points for estimates.

SI Figure 2.3. REVIGO representation of Gene Ontology term enrichment for *Daphnia* swimming behaviors. (Digital upload)



SI Figure 2.4. The chambers used for assaying phototaxis (A) and micro-scale swimming behaviors (B).

SI Tables

SI Table 2.1. Clone-wise summary statistics of *Daphnia* swimming behavior traits; mean value above and standard deviation below. (Digital upload)

SI Table 2.2. *Daphnia* swimming behavior trait-gene associations. (Digital upload)

Global module	night_PI_ctl	night_PI_delta	night_PI_kairo	delta_area	delta_cv_v	delta_max_v	delta_mean_v	delta_sinuosity	post_area	post_cv_v	post_max_v	post_mean_v	post_med_v	post_sinuosity	pre_area	pre_cv_v	pre_max_v	pre_mean_v	pre_med_v	pre_sinuosity
blue	13	20	58	25	26	14	31	13	32	18	16	19	12	-	12	71	-	14	-	12
green	-	25	-	33	17	18	13	-	39	16	-	13	11	-	16	-	-	-	-	-
black	-	-	-	40	12	23	20	x	81	21	42	44	25	-	16	x	18	-	-	-
salmon	-	-	11	24	-	-	23	-	34	-	40	62	63	-	26	x	33	38	17	-
brown	-	28	33	34	-	19	14	-	32	-	22	16	-	14	-	-	-	-	-	-
purple	33	47	-	14	15	32	16	-	18	-	18	12	38	-	-	13	-	15	32	x
turquoise	27	-	11	14	24	-	-	-	16	17	11	-	-	12	11	-	-	-	-	-
cyan	56	38	-	-	-	-	x	-	-	x	x	-	-	-	x	64	-	-	32	-
magenta	-	-	-	-	x	-	-	-	-	-	x	-	-	105	x	-	-	-	-	97
midnightblue	13	20	-	13	23	-	-	x	11	22	-	-	-	-	-	-	-	x	x	-
grey60	-	x	-	-	x	-	-	x	-	-	-	-	-	-	-	-	x	x	-	-
darkgrey	-	x	-	7	x	7	9	-	13	-	18	17	23	-	-	x	-	-	8	-
yellow	-	-	-	-	-	-	-	-	15	-	-	-	-	-	22	17	13	-	11	-
royalblue	-	-	x	15	-	13	11	-	14	-	22	-	-	-	-	x	-	-	-	-
tan	x	-	-	-	-	-	-	x	-	-	-	12	-	-	-	27	-	-	-	-
lightyellow	-	-	-	-	-	11	x	x	11	-	-	-	-	-	-	x	-	-	-	x
orange	-	-	-	-	x	-	-	-	7	x	8	7	11	24	-	-	-	-	-	-
greenyellow	13	-	-	-	-	x	-	x	-	-	x	13	15	-	-	-	-	-	-	-
pink	x	-	-	-	-	-	-	-	-	-	12	-	-	-	-	-	16	-	-	x
red	-	-	-	11	-	-	-	-	12	-	-	-	-	-	-	-	-	-	-	-
lightgreen	-	-	33	-	-	x	-	-	-	-	x	-	x	-	x	-	-	x	-	-
white	-	-	6	x	x	x	-	x	-	-	x	x	x	-	x	x	-	-	-	x
darkturquoise	-	-	-	-	-	-	x	x	-	-	-	-	12	-	-	x	-	x	x	-
lightcyan	-	-	11	-	x	-	-	x	-	x	-	-	-	x	-	-	-	-	-	-
darkgreen	-	-	-	-	-	x	-	x	-	x	-	-	x	-	-	x	x	x	-	x
darkorange	-	-	x	-	-	x	x	-	-	-	-	-	x	-	-	-	-	-	-	x
darkred	-	x	-	-	x	-	-	-	-	-	-	x	-	x	x	x	-	-	-	-
grey	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-
# global modules	6	6	7	11	6	8	8	1	14	5	10	10	9	4	6	5	4	3	5	2
# trait modules	9	8	4	3	9	9	7	2	7	6	12	9	7	1	3	8	8	8	5	2

SI Table 2.3. Relationship between *Daphnia* swimming trait-associated genes and global coexpression modules.

SI Table 2.4. Gene Ontology (GO) term enrichment of genes associated with *Daphnia* swimming behaviors. (Digital upload)

SI Table 2.5. Promoter motif enrichment of genes associated with *Daphnia* swimming behaviors. (Digital upload)

Trait	Marker	*p*-value
delta_area	scaffold_20.1313161	1.95E-05
delta_area	scaffold_40.110551	5.83E-05
delta_area	scaffold_422.17433	2.42E-05
delta_mean_v	scaffold_10.884591	7.94E-05
delta_mean_v	scaffold_202.154465	2.70E-05
delta_mean_v	scaffold_3.2180642	7.94E-05
night_PI_ctl	scaffold_40.251269	8.03E-06
post_area	scaffold_40.110551	2.85E-05
post_area	scaffold_42.919078	1.05E-05
post_area	scaffold_422.17433	7.41E-06
post_area	scaffold_87.153922	4.01E-05
post_area	scaffold_87.265160	6.66E-05
post_cv_v	scaffold_20.1313161	8.79E-05
post_max_v	scaffold_10.880459	2.03E-05
post_max_v	scaffold_10.884591	6.29E-05
post_max_v	scaffold_100.177784	2.34E-05
post_max_v	scaffold_11.1885366	4.48E-05
post_max_v	scaffold_110.47835	2.34E-05
post_max_v	scaffold_12.756887	1.99E-05
post_max_v	scaffold_18.442525	4.89E-06
post_max_v	scaffold_18.568800	4.70E-05
post_max_v	scaffold_202.154460	9.07E-05
post_max_v	scaffold_3.2180642	6.29E-05
post_max_v	scaffold_36.936785	6.66E-05
post_max_v	scaffold_4.1427693	5.34E-06
post_max_v	scaffold_51.74756	9.56E-05
post_max_v	scaffold_51.84253	1.30E-05
post_max_v	scaffold_7.1472763	5.66E-05
post_max_v	scaffold_7.1614980	8.09E-05
post_max_v	scaffold_79.361481	6.21E-05
post_max_v	scaffold_99.207679	1.01E-05
post_mean_v	scaffold_116.365624	4.28E-05
post_mean_v	scaffold_299.57072	9.64E-05
post_mean_v	scaffold_78.171419	7.05E-05
post_mean_v	scaffold_99.207679	6.94E-05
post_sinuosity	scaffold_61.202646	9.66E-05
pre_max_v	scaffold_13.636806	7.68E-05
pre_max_v	scaffold_67.703935	5.97E-06
pre_sinuosity	scaffold_17.1164223	1.17E-05

SI Table 2.6. SNPs associated with *Daphnia* swimming behaviors at a nominal $p < 1e^{-5}$.

SI Table 2.7. SNPs and differentially expressed genes of an integrated network associated with *Daphnia* swimming behavior trait variation. (Digital upload)

SI Text 1

The genetic structure of populations can inform our understanding of evolutionary history, but creates problems when attempting to determine whether, and which, functional relationships exist between allelic variants and phenotypic variation because history can be confounded with function. Daphnia of the group studied here have been considered by ecologists to consist of two parental species, D. pulex and D. pulicaria, that exist in different habitats but interbreed to produce hybrids that exist in a third habitat. Field surveys have historically found the hybrids to be obligate parthenogens, thus limiting gene flow between the parental species. This history sets up the expectation of strong genetic structure in the panel because we collected clones from all three habitats.

After using GATK to genotype the panel, we converted the matrix to 0/1/2 (representing the number of reference alleles possessed by each clone) for all loci with two alleles (n = 5722). We next used principal components analysis (PCA) implemented in the FactoMineR package (Husson et al. 2013) to decompose the normalized (following Price et al. 2006) genotype matrix. Contrary to expectations, there was very little structure in the panel (Figure SI_T1_1): the first three components each account for just over 4% of total variance. For comparison, we generated a 32 x 5000 matrix of random values on the interval (0, 2), then used the same procedure to decompose the random matrix (Figure SI_T1_2). The first three components of the random decomposition each account for just over 3% each. That is, the genotype data has just over 1% more structure than completely random data in the first three axes; the sum of the absolute values of observed minus random marginal variances across all 32 axes (or total deviation from random) was 8.9%.

We next considered the biplot of PCs 1 and 2 (Figure SI_T1_3). Three general vectors are visible, but two factors reinforce the interpretation of low panel structure. First, the clones along these three general vectors do not correspond to three Daphnia ecotypes: the upper-right quadrat includes pulex (C30), pulicaria (C2), and hybrids (C11); the upper-left, pulex (C21 and C22) and hybrid (C33); and lower-left pulex (C27, C29) and hybrids (e.g., C3). Second, these two axes are a very small portion of the total variance (2.2% departure from random, or 9.3% total).

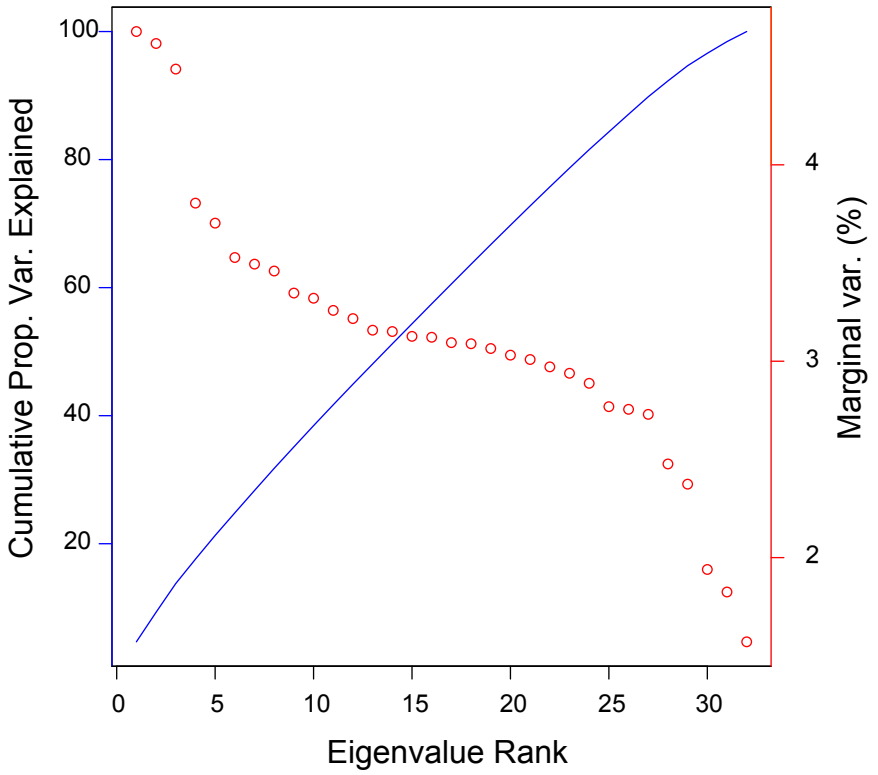
We also analyzed the genotype matrix using STRUCTURE (Pritchard et al. 2000) to determine if model-based analysis provided a different result, given the expectation of population structure. We first removed markers < 200bp distant to reduce LD, then used 10000 burnin replicates and 20000 MCMC reps thereafter to estimate parameters; repeated runs at different K values showed these estimates to be stable. With analyses of $1 \leq K \leq 12$ we found little population structure: the assignment of clones to origin from the K populations were evenly split as 1/K.

Because low minor allele frequency (MAF) can lead to ascertainment bias, we first reduced the matrix to loci with a minor allele frequency > 0.2, and two populations were recovered (posterior Pr(D|K=2) = 0.96). However, these assignments did not match the clone ecotypes: pulicaria, pulex, and hybrids clustered into one population and hybrids and pulex in another. Next, rather than reducing the matrix by enforcing a MAF cutoff, we set lambda, the Dirichlet distribution parameter for random allele draws, to 0.5 following the recommendations of Pritchard et al. (2012) and estimated alpha at $1 \leq K \leq 5$. Under these parameters the posterior Pr(D|K=4) = 0.98, but the population assignments remained ~1/K (Table 1), indicating that although there are four origin populations there is effectively very little structure to the panel. Finally, we performed an

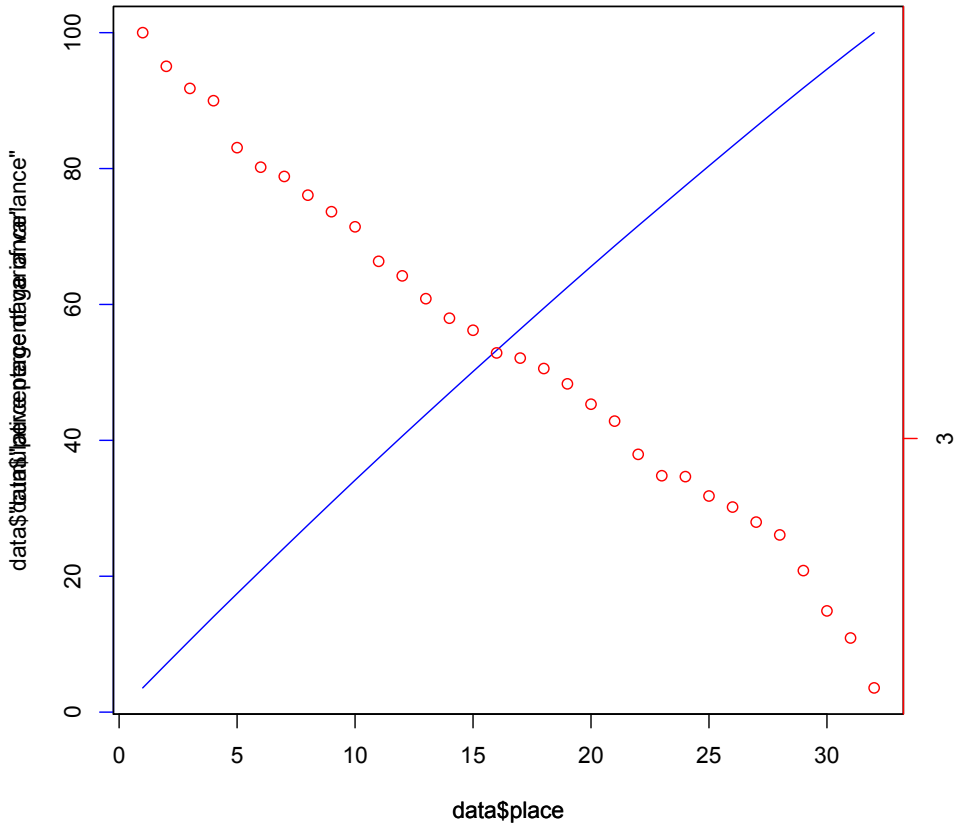
analysis in which we estimated lambda given K=1, then set lambda at the estimate (9.7) for $1 \leq K \leq 5$. Again, there was very little population structure detected, with K=1 possessing the highest –log(likelihood) (= -123155) followed by K=2 (= -123161). Although the K=1 solution is not overwhelmingly supported compared to K=2, as with the high-MAF analysis, the K=2 assignments did not match ecotypes.

Rather than observing population structure along a pulex-pulicaria axis with hybrids in-between, as expected from classical ecological work, we found only slight structure that did not correspond with the three ecotypes. We discuss in the main text (Discussion) other recent breeding and molecular work that supports the presence of gene flow among Daphnia pulex group ecotypes. In addition, we discuss the caveats of these results: this is a relatively small panel from a small geographic area, and extending the results to the Daphnia pulex group in general requires caution (i.e., should be put on hold until larger sample size is acquired).

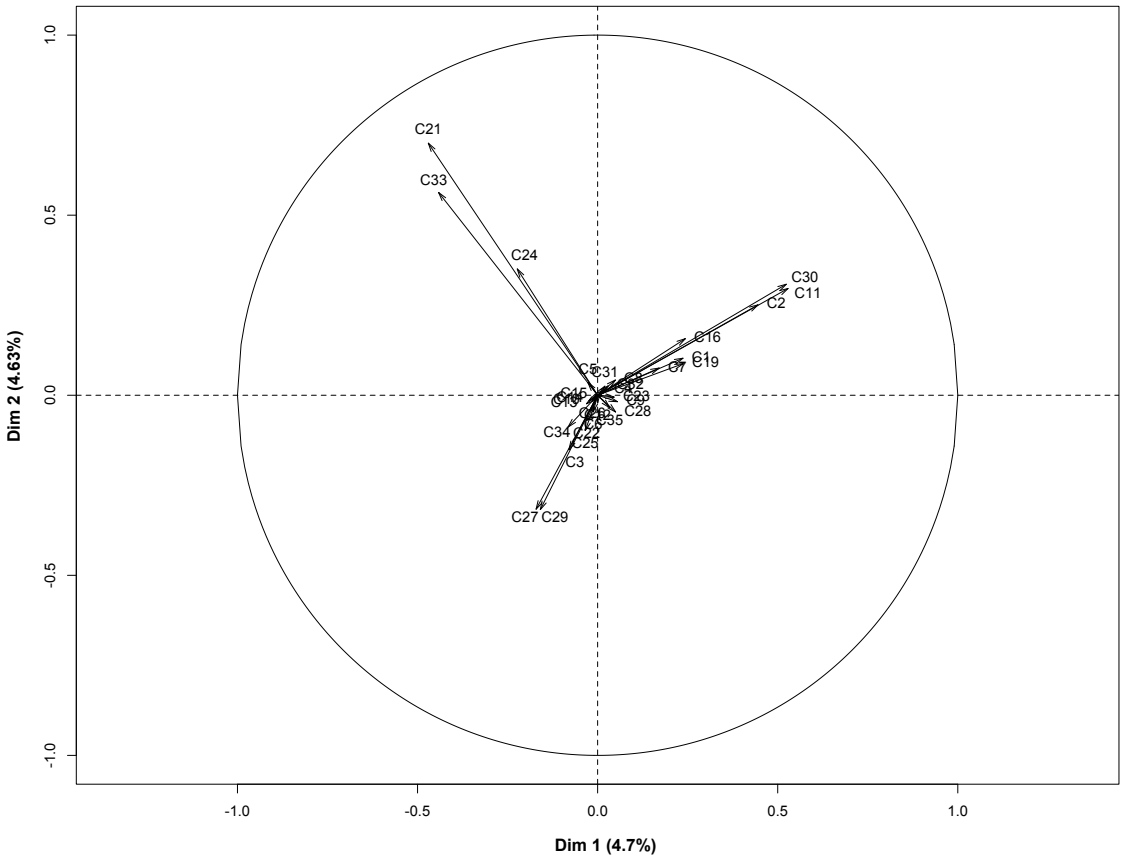
This suggests that, at least with the clones from this panel, there has been recent gene flow among the three ecotypes. Furthermore, the lack of structure enables genome-wide association analyses with reduced concern for structure-function confounding.



SI Text 1 Figure 1. Results of genotype matrix decomposition for the *Daphnia* panel. The blue line and left y-axis give the cumulative proportion of variance explained, while the red points and right y-axis provide the marginal variance of each eigenvector.



SI Text Figure 2. Results of genotype matrix decomposition for a random matrix of approximately the same size (32x5000) as the *Daphnia* genotype matrix. The blue line and left y-axis give the cumulative proportion of variance explained, while the red points and right y-axis provide the marginal variance of each eigenvector.



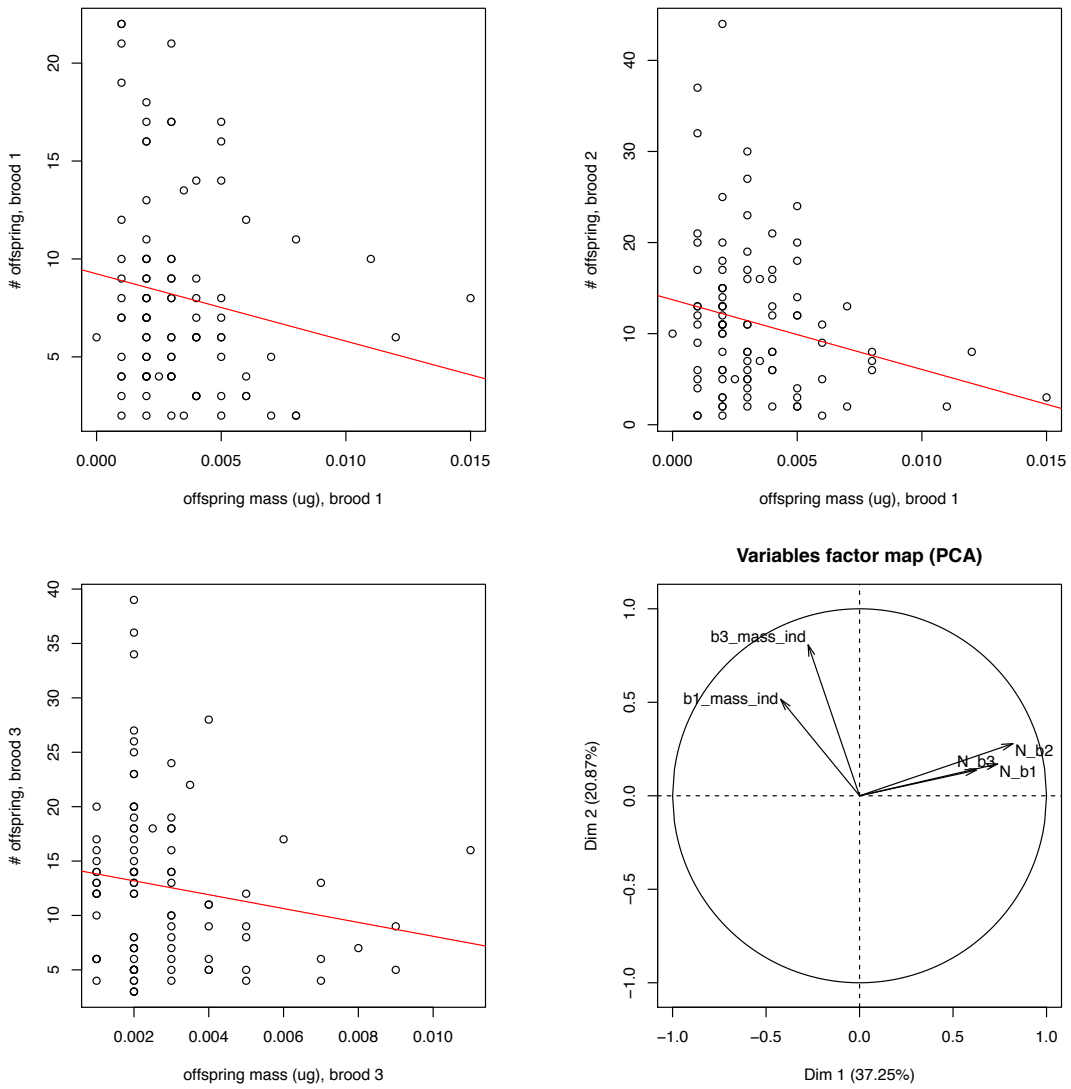
SI Text Figure 3. Biplot of the first two components of PCA decomposition of the Daphnia genotype matrix.

Clone	Pop. 1	Pop. 2	Pop. 3	Pop. 4
C1	0.253	0.248	0.251	0.248
C10	0.247	0.249	0.247	0.256
C11	0.248	0.251	0.251	0.251
C12	0.252	0.25	0.249	0.249
C13	0.251	0.251	0.25	0.248
C14	0.25	0.247	0.252	0.25
C15	0.248	0.25	0.251	0.251
C16	0.25	0.25	0.248	0.253
C19	0.25	0.255	0.248	0.247
C2	0.248	0.253	0.246	0.253
C21	0.253	0.251	0.246	0.25
C22	0.248	0.25	0.248	0.254
C23	0.248	0.253	0.245	0.254
C24	0.256	0.251	0.245	0.248
C25	0.249	0.253	0.249	0.248
C26	0.251	0.251	0.248	0.249
C27	0.253	0.247	0.249	0.25
C28	0.249	0.253	0.251	0.248
C29	0.249	0.248	0.253	0.25
C3	0.249	0.248	0.251	0.253
C30	0.251	0.25	0.247	0.252
C31	0.249	0.255	0.247	0.249
C32	0.249	0.251	0.251	0.249
C33	0.248	0.25	0.249	0.252
C35	0.249	0.254	0.248	0.249
C4	0.248	0.25	0.25	0.252
C5	0.25	0.249	0.249	0.252
C6	0.255	0.248	0.25	0.247
C7	0.249	0.251	0.248	0.253
C8	0.251	0.248	0.249	0.252
C9	0.251	0.251	0.249	0.248
Fst	0.0087	0.0087	0.012	0.0089

SI Text Table 1. Summary of assignments from STRUCTURE analysis for K=4, lambda=0.5.

CHAPTER 3 SUPPLEMENTAL INFORMATION

SI Figures



SI Figure 3.1. Scatterplots of *Daphnia* offspring size and number phenotypic values, and PCA biplot of the phenotypic values.

SI Tables

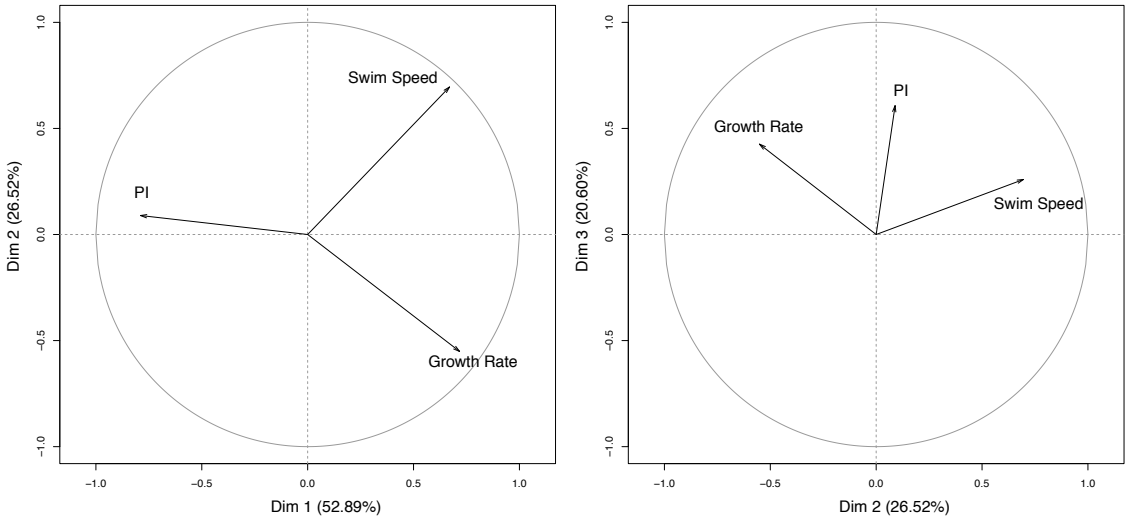
SI Table 3.1. Genes associated with variation in *Daphnia* offspring size-number trade-off. (Digital upload)

SI Table 3.2. SNPs and marker-associated genes (i.e., within 1kb of the SNP) that covary with *Daphnia* offspring size-number trade-off variation. (Digital upload)

SI Table 3.3. Gene Ontology (GO) term enrichment of gene sets associated with *Daphnia* offspring size-number trade-off variation. (Digital upload)

CHAPTER 4 SUPPLEMENTAL INFORMATION

SI Figures



SI Figure 4.1. PCA biplots of the three "keystone traits" of *Daphnia*.

SI Tables

SI Table 4.1. Genes of the 34-gene keystone Candidate Network 1. (Digital upload)

SI Table 4.2. Genes of the 135-gene keystone Candidate Network 2. (Digital upload)

Module	GO domain	GO ID	*p*-value	Odds Ratio	Expect.	Obs.	Size	Term
Blue	MF	GO:0045182	0.0134	88.258	0.014	1	17	translation regulator activity
	BP	GO:0006629	0.0512	6.477	0.373	2	478	lipid metabolic process
	BP	GO:0048869	0.0778	5.061	0.472	2	605	cellular developmental process
	BP	GO:0030154	0.0778	5.061	0.472	2	605	cell differentiation
	BP	GO:0032502	0.0814	3.140	1.753	4	2247	developmental process
	CC	GO:0005576	0.0608	4.238	0.951	3	1242	extracellular region
Brown	CC	GO:0032991	0.0312	5.468	1.374	4	2306	macromolecular complex
	CC	GO:0043234	0.0817	4.052	1.094	3	1837	protein complex
Green	MF	GO:0016773	0.0249	79.284	0.025	1	142	phosphotransferase activity, alcohol group as acceptor
	MF	GO:0004672	0.0249	79.284	0.025	1	142	protein kinase activity
	MF	GO:0016772	0.0482	40.164	0.049	1	276	transferase activity, transferring phosphorus-containing groups
	MF	GO:0016301	0.0482	40.164	0.049	1	276	kinase activity
	MF	GO:0016787	0.0604	4.068	0.492	2	2783	hydrolase activity
	CC	GO:0031012	0.0491	26.484	0.050	1	147	extracellular matrix
	CC	GO:0005578	0.0491	26.484	0.050	1	147	proteinaceous extracellular matrix
Grey	BP	GO:0043412	0.0463	3.674	1.440	4	1538	macromolecule modification
	BP	GO:0036211	0.0463	3.674	1.440	4	1538	protein modification process
	BP	GO:0006464	0.0463	3.674	1.440	4	1538	cellular protein modification process
	BP	GO:0044267	0.0695	3.165	1.639	4	1751	cellular protein metabolic process
	BP	GO:0007275	0.0868	3.372	1.081	3	1155	multicellular organismal development
	BP	GO:0009607	0.0897	11.669	0.094	1	100	response to biotic stimulus
	BP	GO:0007010	0.0944	4.382	0.525	2	561	cytoskeleton organization
	CC	GO:0005856	0.0130	7.461	0.516	3	505	cytoskeleton
Red	MF	GO:0016209	0.0455	28.687	0.046	1	131	antioxidant activity
	CC	GO:0005783	0.0194	15.925	0.237	2	696	endoplasmic reticulum
Turquoise	MF	GO:0008233	0.0085	7.181	0.905	4	1138	peptidase activity
	MF	GO:0016787	0.0457	3.840	2.212	5	2783	hydrolase activity
	MF	GO:0003824	0.0739	3.755	4.344	7	5465	catalytic activity
	CC	GO:0031012	0.0002	40.267	0.113	3	147	extracellular matrix
	CC	GO:0005578	0.0002	40.267	0.113	3	147	proteinaceous extracellular matrix
	CC	GO:0005576	0.0011	10.614	0.951	5	1242	extracellular region
	CC	GO:0044421	0.0041	12.459	0.349	3	456	extracellular region part
Yellow	BP	GO:0044238	0.0790	6.280	1.946	4	4990	primary metabolic process

SI Table 4.3. Module-wise Gene Ontology (GO) term enrichment of the genes of Candidate Network 2.

Bibliography

Abrams, P.A. & Rowe, L. (1996). The effects of predation on the age and size of maturity of prey. Evolution, 1052-1061.

Acharya, K., Kyle, M. & Elser, J.J. (2004). Effects of stoichiometric dietary mixing on *Daphnia* growth and reproduction. Oecologia, 138, 333-340.

Ackerly, D.D. & Cornwell, W.K. (2007). A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. Ecol. Lett., 10, 135-145.

Adamowicz, S.J., Petrusek, A., Colbourne, J.K., Hebert, P.D.N. & Witt, J.D.S. (2009). The scale of divergence: A phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. Mol. Phylogenet. Evol., 50, 423-436.

Agar, W.E. (1914). Experiments on Inheritance in Parthenogenesis. Philos. Trans. R. Soc. Lond. Ser. B Contain. Pap. Biol. Character, 205, 421-489.

Alekseev, V. & Lampert, W. (2001). Maternal control of resting-egg production in *Daphnia*. Nature, 414, 899-901.

Alon, U. (2003). Biological networks: the tinkerer as an engineer. Science, 301, 1866-1867.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25, 3389-3402.

Antonovics, J. (2003). Toward community genomics? Ecology, 84, 598-601.

Arendt, J. & Reznick, D. (2008). Convergence and parallelism reconsidered: what have I learned about the genetics of adaptation? Trends Ecol. Evol., 23, 26-32.

Arnold, S.J., Burger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. (2008). Understanding The Evolution And Stability Of The G-Matrix. Evolution, 62, 2451-2461.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet., 25, 25-29.

Asselman, J., De Coninck, D.I.M., Glaholt, S., Colbourne, J.K., Janssen, C.R., Shaw, J.R., et al. (2012). Identification of Pathways, Gene Networks, and Paralogous Gene Families in *Daphnia pulex* Responding to Exposure to the Toxic Cyanobacterium *Microcystis aeruginosa*. Environ. Sci. Technol., 46, 8448-8457.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., et al. (2009). Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet, 41, 299-307.

Barabasi, A.-L. & Oltvai, Z.N. (2004). Network biology: understanding the cell,Äôs functional organization. Nat Rev Genet, 5, 101-113.

Barton, N.H., Briggs, D.E.G., Eisen, J.A., Goldstein, D.B. & Patel, N.H. (2007). Evolution. Cold Springs Harbor Laboratory Press, New York.

Beckerman, A.P., Rodgers, G.M. & Dennis, S.R. (2010). The reaction norm of size and age at maturity under multiple predator risk. J. Anim. Ecol., 79, 1069-1076.

Bell, A.M. (2009). Approaching the Genomics of Risk-Taking Behavior. In: Advances in Genetics. Elsevier, pp. 83-104.

Bernot, R., Dodds, W., Quist, M. & Guy, C. (2006). Temperature and kairomone induced life history plasticity in coexisting *Daphnia*. Aquat. Ecol., 40, 361-372.

Biesecker, L.G. (2013). Hypothesis-generating research and predictive medicine. Genome Res., 23, 1051-1053.

Boeing, W.J., Ramcharan, C.W. & Riessen, H.P. (2006). Multiple predator defence strategies in *Daphnia pulex* and their relation to native habitat. J. Plankton Res., 28, 571 -584.

Boersma, M. (1997). Offspring size and parental fitness in *Daphnia magna*. Evol. Ecol., 11, 439-450.

Boersma, M., Spaak, P. & De Meester, L. (1998). Predator-mediated plasticity in morphology, life history, and behavior of *Daphnia*: the uncoupling of responses. The American Naturalist, 152, 237-248.

Bogdan, K.G. & Gilbert, J.J. (1984). Body size and food size in freshwater zooplankton. Proc. Natl. Acad. Sci., 81, 6427.

Boucher, P., Ditlecadet, D., Dubé, C. & Dufresne, F. (2010). Unusual duplication of the insulin-like receptor in the crustacean *Daphnia pulex*. Bmc Evol. Biol., 10, 305.

Bownes, M. (1992). Why is there sequence similarity between insect yolk proteins and vertebrate lipases? J. Lipid Res., 33, 777-790.

Bownes, M., Lineruth, K. & Mauchline, D. (1991). Egg production and fertility in *Drosophila* depend upon the number of yolk-protein gene copies. Mol. Gen. Genet. Mgg, 228, 324-327.

Bresnick, A.R. (1999). Molecular mechanisms of nonmuscle myosin-II regulation. Current Opinion in Cell Biology, 11, 26-33.

Brewer, M.C., Dawidowicz, P. & Dodson, S.I. (1999). Interactive effects of fish kairomone and light on *Daphnia* escape behavior. *J. Plank. Res*., 21, 1317-1335.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R. & Hafen, E. (2001). An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. Curr. Biol., 11, 213-221.

Bronstein, R., Levkovitz, L., Yosef, N., Yanku, M., Ruppin, E., Sharan, R., et al. (2010). Transcriptional Regulation by CHIP/LDB Complexes. PLoS Genet, 6, e1001063.

Brooks, J.L. & Dodson, S.I. (1965). Predation, Body Size, and Composition of Plankton. Science, 150, 28.

Bruggeman, F.J. & Westerhoff, H.V. (2007). The nature of systems biology. Trends Microbiol., 15, 45-50.

Carhan, A., Reeve, S., Dee, C.T., Baines, R.A. & Moffat, K.G. (2004). Mutation in slowmo causes defects in *Drosophila* larval locomotor behaviour. Invert. Neurosci., 5, 65-75.

Carpenter, S.R. & Kitchell, J.F. (1996). The trophic cascade in lakes. Cambridge University Press.

Carroll, S., Hendry, A., Reznick, D. & Fox, C. (2007). Evolution on ecological time-scales. Funct. Ecol., 21, 387-393.

Carroll, S.B. (2009). Endless forms: the evolution of gene regulation and morphological diversity. Philosophy of Biology: An Anthology, 32, 193.

Cerny, M. & Bytel, J. (1991). Density and size distribution of *Daphnia* populations at different fish predation levels. Hydrobiologia, 225, 199-208.

Chamberlin, T.C. (1890). The method of multiple working hypotheses. Science, 15, 92-96.

Chase, J.M. & Leibold, M.A. (2003). Ecological Niches: Linking Classical and Contemporary Approaches. University Of Chicago Press.

Christie, A.E. & McCoole, M.D. (2012). From genes to behavior: investigations of neurochemical signaling come of age for the model crustacean *Daphnia pulex*. J Exp Biol, 215, 2535-2544.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., et al. (2011). The Ecoresponsive Genome of *Daphnia pulex*. Science, 331, 555 -561.

Cooper, S.D. & Smith, D.W. (1982). Competition, predation and the relative abundances of two species of *Daphnia*. J. Plankton Res., 4, 859-879.

Cousyn, C., De Meester, L., Colbourne, J.K., Brendonck, L., Verschuren, D. & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. Proc. Natl. Acad. Sci., 98, 6256.

Cristescu, M.E., Constantin, A., Bock, D.G., Cáceres, C.E. & Crease, T.J. (2012). Speciation with gene flow and the genetics of habitat transitions. Mol. Ecol., 21, 1411-1422.

Cristescu, M.E.A., Colbourne, J.K., Radivojac, J. & Lynch, M. (2006). A microsatellite-based genetic linkage map of the waterflea, *Daphnia pulex*: On the prospect of crustacean genomics. Genomics, 88, 415-430.

David, M., Dzamba, M., Lister, D., Ilie, L. & Brudno, M. (2011). SHRiMP2: Sensitive yet Practical Short Read Mapping. Bioinformatics, 27, 1011 -1012.

Dawidowicz, P. & Loose, C.J. (1992). Cost of Swimming by *Daphnia* During Diel Vertical Migration. Limnology and Oceanography, 37, 665-669.

Dawkins, R. (1982). The Extended Phenotype: The Long Reach of the Gene. Oxford University Press, USA.

De Mazancourt, C. & Dieckmann, U. (2004). Trade-Off Geometries and Frequency-Dependent Selection. Am. Nat., 164, 765-778.

De Meester, L. & Cousyn, C. (1997). The change in phototactic behaviour of a *Daphnia magna* clone in the presence of fish kairomones: the effect of exposure time. Hydrobiologia, 360, 169-175.

De Meester, L. & Weider, L.J. (1999). Depth selection behavior, fish kairomones, and the life histories of *Daphnia* hyalina X galeata hybrid clones. Limnol. Ocean., 1248-1258.

De Meester, L. (1991). An analysis of the phototactic behaviour of *Daphnia magna* clones and their sexual descendants. Hydrobiologia, 225, 217-227.

De Meester, L. (1993). Genotype, fish-mediated chemical, and phototactic behavior in *Daphnia magna*. Ecology, 74, 1467-1474.

De Meester, L. (1996a). Evolutionary potential and local genetic differentiation in a phenotypically plastic trait of a cyclical parthenogen, *Daphnia magna*. Evolution, 1293-1298.

De Meester, L., Gomez, A., Okamura, B. & Schwenk, K. (2002). The Monopolization Hypothesis and the dispersal-gene flow paradox in aquatic organisms. Acta Oecologica, 23, 121-135.

DeMott, W.R. & Kerfoot, W.C. (1982). Competition among cladocerans: nature of the interaction between Bosmina and *Daphnia*. Ecology, 63, 1949-1966.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet., 43, 491-498.

Dennis, S.R., Carter, M.J., Hentley, W.T. & Beckerman, A.P. (2011). Phenotypic convergence along a gradient of predation risk. Proc. R. Soc. B Biol. Sci., 278, 1687 -1696.

Dini, M.L. & Carpenter, S.R. (1992). Fish predators, food availability and diel vertical migration in *Daphnia*. J. Plankton Res., 14, 359-377.

Dodson, S. & Ramcharan, C. (1991). Size-specific swimming behavior of *Daphnia pulex*. J. Plankton Res., 13, 1367 -1379.

Dodson, S. (1988). The Ecological Role of Chemical Stimuli for the Zooplankton: Predator-Avoidance Behavior in *Daphnia*. Limnology and Oceanography, 33, 1431-1439.

Dodson, S.I. & Hanazato, T. (1995). Commentary on effects of anthropogenic and natural organic chemicals on development, swimming behavior, and reproduction of *Daphnia*, a key member of aquatic ecosystems. Env. Health Persp., 103, 7.

Dodson, S.I. (1974). Zooplankton Competition and Predation: An Experimental Test of the Size-Efficiency Hypothesis. Ecology, 55, 605-613.

Dodson, S.I., Newman, A.L., Will-Wolf, S., Alexander, M.L., Woodford, M.P. & Van Egeren, S. (2009). The relationship between zooplankton community structure and lake characteristics in temperate lakes (Northern Wisconsin, USA). J. Plankton Res., 31, 93-100.

Doksaeter, A. & Vijverberg, J. (2001). The effects of food and temperature regimes on life-history responses to fish kairomones in *Daphnia* hyalina X galeata*. Hydrobiologia, 442, 207-214.

Drenner, R.W., Strickler, J.R. & O,ÄôBrien, W.J. (1978). Capture Probability: The Role of Zooplankter Escape in the Selective Feeding of Planktivorous Fish. J. Fish. Res. Bd. Can., 35, 1370-1373.

Dudycha, J.L. & Tessier, A.J. (1999). Natural Genetic Variation of Life Span, Reproduction, and Juvenile Growth in *Daphnia*. Evolution, 53, 1744-1756.

Dudycha, J.L., Brandon, C.S. & Deitz, K.C. (2012). Population genomics of resource exploitation: insights from gene expression profiles of two *Daphnia* ecotypes fed alternate resources. Ecol. Evol.

Duffy, M.A. & Sivars-Becker, L. (2007). Rapid evolution and ecological host-parasite dynamics. Ecol. Lett., 10, 44-53.

Duffy, M.A. (2010). Ecological consequences of intraspecific variation in lake *Daphnia*. Freshwater Biology, 55, 995-1004.

Duffy, M.A., Hall, S.R., Cáceres, C.E. & Ives, A.R. (2009). Rapid evolution, seasonality, and the termination of parasite epidemics. Ecology, 90, 1441-1448.

Ebert, D. (1993). The trade-off between offspring size and number in *Daphnia magna*: the influence of genetic, environmental and maternal effects. Arch. Hydrobiol.-Suppl.-, 90, 453-453.

Einum, S. & Fleming, I.A. (2000). Highly fecund mothers sacrifice offspring survival to maximize fitness. Nature, 405, 565-567.

Ellner, S.P., Geber, M.A. & Hairston Jr, N.G. (2011). Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. Ecol. Lett.

Elser, J.J., Acharya, K., Kyle, M., Cotner, J., *Magna*, W., Markow, T., et al. (2003). Growth rate-stoichiometry couplings in diverse biota. Ecol. Lett., 6, 936-943.

Enquist, B.J., West, G.B., Charnov, E.L. & Brown, J.H. (1999). Allometric scaling of production and life-history variation in vascular plants. Nature, 401, 907-911.

Enserink, E., Maasdiepeveen, J. & Vanleeuwen, C. (1991). Combined effects of metals - an ecotoxicological evaluation. Water Research, 25, 679-687.

Espinosa-Soto, C. & Wagner, A. (2010). Specialization Can Drive the Evolution of Modularity. PLoS Comput Biol, 6, e1000719.

Ewens, W.J. & Grant, G.R. (2010). Statistical Methods in Bioinformatics: An Introduction. Springer.

Falcon, S. & Gentleman, R. (2011). How To Use GOstats Testing Gene Lists for GO Term Association.

Feder, M.E. & Mitchell-Olds, T. (2003). Evolutionary and ecological functional genomics. Nat. Rev. Genet., 4, 649-655.

Fisk, D.L., Latta, L.C., Knapp, R.A. & Pfrender, M.E. (2007). Rapid evolution in response to introduced predators I: rates and patterns of morphological and life-history trait divergence. Bmc Evol. Biol., 7, 22.

Forbes, S.A. (1880). The Food of the Darters. Am. Nat., 14, 697-703.

Fujiwara, M., Sengupta, P. & McIntire, S.L. (2002). Regulation of Body Size and Behavioral State of *C. elegans* by Sensory Perception and the EGL-4 cGMP-Dependent Protein Kinase. Neuron, 36, 1091-1102.

Glazier, D.S. (1992). Effects of Food, Genotype, and Maternal Size and Age on Offspring Investment in *Daphnia Magna*. Ecology, 73, 910-926.

Gliwicz, Z. & Umana, G. (1994). Cladoceran body-size and vulnerability to copepod predation. Limnol. Ocean., 39, 419-424.

Gliwicz, Z. & Wrzosek, D. (2008). Predation-mediated coexistence of large- and small-bodied *Daphnia* at different food levels. Am. Nat., 172, 358-374.

Gliwicz, Z.M. & Maszczyk, P. (2006). *Daphnia* growth is hindered by chemical information on predation risk at high but not at low food levels. Oecologia, 150, 706-715.

Godfray, H.C.J. (1995). Evolutionary theory of parent offspring conflict. Nature, 376, 133-138.

Goulden, C.E., Henry, L.L. & Tessier, A.J. (1982). Body Size, Energy Reserves, and Competitive Ability in Three Species of Cladocera. Ecology, 63, 1780-1789.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol., 8, R24.

Hahn, M., Conant, G. & Wagner, A. (2004). Molecular Evolution in Large Genetic Networks: Does Connectivity Equal Constraint? J. Mol. Evol., 58, 203-211.

Hairston, N., Lampert, W., Caceres, C., Holtmeier, C., Weider, L., Gaedke, U., et al. (1999). Lake ecosystems - Rapid evolution revealed by dormant eggs. Nature, 401, 446-446.

Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T. & Fox, J.A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. Ecol. Lett., 8, 1114-1127.

Hall, S., Leibold, M., Lytle, D. & Smith, V. (2004). Stoichiometry and planktonic grazer composition over gradients of light, nutrients, and predation risk. Ecology, 85, 2291-2301.

Hall, S.R., Sivars-Becker, L., Becker, C., Duffy, M.A., Tessier, A.J. & Cáceres, C.E. (2007). Eating yourself sick: transmission of disease as a function of foraging ecology. Ecology Letters, 10, 207-218.

Hamilton, M. (2009). Population Genetics. 1st edn. Wiley-Blackwell.

Hanazato, T. & Dodson, S.I. (1995). Synergistic effects of low oxygen concentration, predator kairomone, and a pesticide on the cladoceran *Daphnia pulex*. Limnology and Oceanography, 700-709.

Hansen, T. (2006). The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst., 37, 123-157.

Hansen, T.F. (2003). Is modularity necessary for evolvability?: Remarks on the relationship between pleiotropy and evolvability. Biosystems, 69, 83-94.

Hart, R.C. & Bychek, E.A. (2010). Body size in freshwater planktonic crustaceans: an overview of extrinsic determinants and modifying influences of biotic interactions. Hydrobiologia, 668, 61-108.

Hartmann, H., Guthöhrlein, E.W., Siebert, M., Luehr, S. & Söding, J. (2013). P-value-based regulatory motif discovery using positional weight matrices. Genome Res., 23, 181-194.

Hebert, P.D.N. & Crease, T. (1983). Clonal diversity in populations of *Daphnia pulex* reproducing by obligate parthenogenesis. Heredity, 51, 353-369.

Hebert, P.D.N., Ward, R.D. & Weider, L.J. (1988). Clonal-Diversity Patterns and Breeding-System Variation in *Daphnia pulex*, asexual-Sexual Complex. Evolution, 42, 147-159.

Heckmann, L.-H., Sibly, R.M., Connon, R., Hooper, H.L., Hutchinson, T.H., Maund, S.J., et al. (2008). Systems biology meets stress ecology: linking molecular and organismal stress responses in *Daphnia magna*. Genome Biol, 9, R40.

Heier, C. & Dudycha, J.L. (2009). Ecological speciation in a cyclic parthenogen: Sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia pulicaria*. Limnol. Ocean., 54, 492-502.

Hodgins-Davis, A. & Townsend, J.P. (2009). Evolving gene expression: from G to E to G X E. Trends Ecol. Evol., 24, 649-658.

Hoekstra, H.E. & Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution, 61, 995-1016.

Hood, J.M. & Sterner, R.W. (2010). Diet Mixing: Do Animals Integrate Growth or Resources across Temporal Heterogeneity? Am. Nat., 176, 651-663.

Hood, J.M. & Sterner, R.W. (2010). Diet Mixing: Do Animals Integrate Growth or Resources across Temporal Heterogeneity? The American Naturalist, 176, 651-663.

Houle, D. (1991). Genetic Covariance of Fitness Correlates: What Genetic Correlations are Made of and Why it Matters. Evolution, 45, 630-648.

Houle, D. (2009). Evolution in Health and Medicine Sackler Colloquium: Numbering the hairs on myheads: The shared challenge and promise of phenomics. Proc. Natl. Acad. Sci.

Houle, D., Govindaraju, D.R. & Omholt, S. (2010). Phenomics: the next challenge. Nat Rev Genet, 11, 855-866.

Hulsmann, S., Rinke, K. & Mooij, W.M. (2011). Size-selective predation and predator-induced life-history shifts alter the outcome of competition between planktonic grazers. Functional Ecology, 25, 199-208.

Ideker, T. & Krogan, N.J. (2012). Differential network biology. Mol Syst Biol, 8.

Inoue, K. & Lupski, J.R. (2003). Genetics and genomics of behavioral and psychiatric disorders. Current opinion in genetics & development, 13, 303-309.

Ishimoto, H. & Kitamoto, T. (2010). The Steroid Molting Hormone Ecdysone Regulates Sleep in Adult *Drosophila melanogaster*. Genetics, 185, 269-281.

Jean-Christophe, S., Michael, P.E., Ralph, T., Denis, T. & John, C.K. (2011). Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. J. Hered.

Jeyasingh, P.D., Ragavendran, A., Paland, S., Lopez, J.A., Sterner, R.W. & Colbourne, J.K. (2011). How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*. Mol. Ecol., 20, 2341-2352.

Ji, Y., Xu, Y., Zhang, Q., Tsui, K.-W., Yuan, Y., Norris Jr., C., et al. (2011). BM-Map: Bayesian Mapping of Multireads for Next-Generation Sequencing Data. Biometrics, 67, 1215-1224.

Jordan, K., Carbone, M., Yamamoto, A., Morgan, T. & Mackay, T. (2007). Quantitative genomics of locomotor behavior in *Drosophila melanogaster*. Genome Biology, 8, R172.

Jumbo-Lucioni, P., Ayroles, J., Chambers, M., Jordan, K., Leips, J., Mackay, T., et al. (2010). Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*. Bmc Genomics, 11, 297.

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 28, 27-30.

Kato, Y., Matsuura, T. & Watanabe, H. (2012). Genomic Integration and Germline Transmission of Plasmid Injected into Crustacean *Daphnia magna* Eggs. Plos One, 7, e45318.

Kato, Y., Shiga, Y., Kobayashi, K., Tokishita, S., Yamagata, H., Iguchi, T., et al. (2011). Development of an RNA interference method in the cladoceran crustacean *Daphnia magna*. Dev Genes Evol.

Kerfoot, W.C., Kellogg Jr, D.L. & Strickler, J.R. (1980). Visual observations of live zooplankters: evasion, escape, and chemical defenses. Evol. Ecol. Zooplankton Communities Univ. Press New Engl., 10-27.

Kirkpatrick, M. (2009). Patterns of quantitative genetic variation in multiple dimensions. Genetica, 136, 271-284.

Klüttgen, B., Dülmer, U., Engels, M. & Ratte, H.. (1994). ADaM, an artificial freshwater for the culture of zooplankton. Water Res., 28, 743-746.

Knight, J.C. (2004). Allele-specific gene expression uncovered. Trends Genet., 20, 113-116.

Kolm, N., Goodwin, N.B., Balshine, S. & Reynolds, J.D. (2006). Life history evolution in cichlids 2: directional evolution of the trade-off between egg number and egg size. J. Evol. Biol., 19, 76-84.

Kozlowski, J. & Stearns, S.C. (1989). Hypotheses for the Production of Excess Zygotes: Models of Bet-Hedging and Selective Abortion. Evolution, 43, 1369-1377.

Kreutzer, C. & Lampert, W. (1999). Exploitative competition in differently sized *Daphnia* species: a mechanistic explanation. Ecology, 80, 2348-2357.

Lampert, W. & Taylor, B.E. (1985). Zooplankton Grazing in a Eutrophic Lake: Implications of Diel Vertical Migration. Ecology, 66, 68-82.

Lampert, W. (1989). The Adaptive Significance of Diel Vertical Migration of Zooplankton. Funct. Ecol., 3, 21-27.

Lampert, W. (1993). Phenotypic Plasticity of the Size at First Reproduction in *Daphnia*: The Importance of Maternal Size. Ecology, 74, 1455-1466.

Lampert, W. (2006). *Daphnia*: model herbivore, predator and prey. Pol. J. Ecol., 54, 607-620.

Lande, R. & Arnold, S.J. (1983). The measurement of selection on correlated characters. Evolution, 1210-1226.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution, 402-416.

Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

Lee, K.-S., You, K.-H., Choo, J.-K., Han, Y.-M. & Yu, K. (2004). *Drosophila* Short Neuropeptide F Regulates Food Intake and Body Size. J. Biol. Chem., 279, 50781-50789.

Leibold, M. & Tessier, A.J. (1991). Contrasting patterns of body size for *Daphnia* species that segregate by habitat. Oecologia, 86, 342-348.

Leibold, M., Tessier, A. & West, C. (1994). Genetic, acclimatization, and ontogenic effects on habitat selection behavior in *Daphnia pulicaria*. Evolution, 48, 1324-1332.

Leibold, M.A. (1989). Resource edibility and the effects of predators and productivity on the outcome of trophic interactions. Am Nat, 134, 922.

Leibold, M.A. (1991). Trophic interactions and habitat segregation between competing *Daphnia* species. Oecologia, 86, 510-520.

Leibold, M.A. (1996). A graphical model of keystone predators in food webs: Trophic regulation of abundance, incidence, and diversity patterns in communities. Am. Nat., 147, 784-812.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009a). The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078-2079.

Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., LeProust, E., et al. (2009b). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science, 324, 1210-1213.

Loewe, L. (2009). A framework for evolutionary systems biology. Bmc Syst. Biol., 3, 27.

Loose, C.J. & Dawidowicz, P. (1994). Trade-offs in diel vertical migration by zooplankton: the costs of predator avoidance. Ecology, 2255-2263.

Loose, C.J., Von Elert, E. & Dawidowicz, P. (1993). Chemically-induced diel vertical migration in *Daphnia*: a new bioassay for kairomones exuded by fish. Archiv fur Hydrobiologie. Stuttgart, 126, 329-337.

Loreau, M. (2010). From Populations to Ecosystems: Theoretical Foundations for a New Ecological Synthesis (MPB-46). Princeton University Press.

Lovell, J.T., Juenger, T.E., Michaels, S.D., Lasky, J.R., Platt, A., Richards, J.H., et al. (2013). Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. Proc. R. Soc. B Biol. Sci., 280.

Lynch, M. & Walsh, B. (1998). Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA. 980 pp.

Lynch, M. (1980). The Evolution of Cladoceran Life Histories. Q. Rev. Biol., 55, 23-42.

Lynch, M. (1984). The limits to life history evolution in *Daphnia*. Evolution, 465-482.

Lynch, M., Latta, L., Hicks, J. & Giorgianni, M. (1998). Mutation, Selection, and the Maintenance of Life-History Variation in a Natural Population. Evolution, 52, 727-733.

Lopez-Maury, L., Marguerat, S. & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat. Rev. Genet., 9, 583-593.

Mackay, T.F.C. (2001). The genetic architecture of quantitative traits. Annu. Rev. Genet., 35, 303-339.

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. Nature, 482, 173-178.

Mackay, T.F.C., Stone, E.A. & Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. Nat Rev Genet, 10, 565-577.

Main, T., Dobberfuhl, D. & Elser, J. (1997). N : P stoichiometry and ontogeny of crustacean zooplankton: A test of the growth rate hypothesis. Limnol. Ocean., 42, 1474-1478.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Candidate loci, genes, and networks underlying ecologically important behavioral variation in a model ecological species, *Daphnia pulex*.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). Exploring the molecular basis of variation in size and fecundity in an emerging model organism, *Daphnia pulex*.

Malcom, J.W., Juenger, T.E. & Leibold, M.A. (in prep.). The joys of emerging model genomic species: Evolutionary patterns and molecular biological function in *Daphnia*.

Mappes, T. & Koskela, E. (2004). Genetic Basis of the Trade-Off Between Offspring Number and Quality in the Bank Vole. Evolution, 58, 645-650.

McCoole, M.D., Baer, K.N. & Christie, A.E. (2011). Histaminergic signaling in the central nervous system of *Daphnia* and a role for it in the control of phototactic behavior. The Journal of Experimental Biology, 214, 1773 -1782.

McCullagh, P. & Nelder, J.A. (1999). Generalized linear models. Chapman & Hall, CRC.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. Trends Ecol. Evol., 21, 178-185.

McGuffin, P., Riley, B. & Plomin, R. (2001). Toward behavioral genomics. Science, 291, 1232-1249.

Magna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res., 20, 1297-1303.

Mendive, F.M., Van Loy, T., Claeysen, S., Poels, J., Williamson, M., Hauser, F., et al. (2005). *Drosophila* molting neurohormone bursicon is a heterodimer and the natural agonist of the orphan receptor DLGR2. Febs Lett., 579, 2171-2176.

Messier, J., McGill, B.J. & Lechowicz, M.J. (2010). How do traits vary across ecological scales? A case for trait-based ecology. Ecol. Lett., 9999.

Meyer, E., Aglyamova, G.V. & Matz, M.V. (2011). Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. Mol. Ecol., 20, 3599-3616.

Mikulski, A. & Pijanowska, J. (2010). When and how can *Daphnia* prepare their offspring for the threat of predation? Hydrobiologia, 643, 21-26.

Mills, L.S., Soulé, M.E. & Doak, D.F. (1993). The keystone-species concept in ecology and conservation. BioScience, 43, 219-224.

Miner, B.E., Meester, L.D., Pfrender, M.E., Lampert, W. & Hairston, N.G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proc. R. Soc. B, 279, 1873-1882.

Mitchell, S. & Read, A. (2005). Poor maternal environment enhances offspring disease resistance in an invertebrate. Proc. R. Soc. B-Biol. Sci., 272, 2601-2607.

Moon, S.Y. & Zheng, Y. (2003). Rho GTPase-activating proteins in cell regulation. Trends in Cell Biology, 13, 13-22.

Nadeau, J.H. & Dudley, A.M. (2011). Systems Genetics. Science, 331, 1015-1016.

Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., et al. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics, 30, 190-193.

Odling-Smee, F.J., Laland, K.N. & Feldman, M.W. (2003). Niche Construction: The Neglected Process in Evolution (MPB-37). Princeton University Press.

Ogasawara, O. & Okubo, K. (2009). On Theoretical Models of Gene Expression Evolution with Random Genetic Drift and Natural Selection. PLoS ONE, 4, e7943.

Ohgushi, T. (1996). A reproductive tradeoff in an herbivorous lady beetle: egg resorption and female survival. Oecologia, 106, 345-351.

O'Keefe, T.C., Brewer, M.C. & Dodson, S.I. (1998). Swimming behavior of *Daphnia*: its role in determining predation risk. *J. Plank. Res.*, 20, 973-984.

Paine, R.T. (1969). A note on trophic complexity and community stability. Am. Nat., 103, 91-93.

Pantel, J.H., Juenger, T.E. & Leibold, M.A. (2011). Environmental gradients structure *Daphnia pulex* X *pulicaria* clonal distribution. J. Evol. Biol., 24, 723-732.

Pantel, J.H., Leibold, M.A. & Juenger, T.E. (2011). Population differentiation in *Daphnia* alters community assembly in experimental ponds. Am. Nat., 177, 314-322.

Pauwels, K., Stoks, R. & De Meester, L. (2005). Coping with predator stress: interclonal differences in induction of heat-shock proteins in the water flea *Daphnia magna*. Journal of evolutionary biology, 18, 867-872.

Pease, C.M. & Bull, J.J. (1988). A critique of methods for measuring life history trade-offs. J. Evol. Biol., 1, 293-303.

Penalva-Arana, D.C., Lynch, M. & Robertson, H. (2009). The chemoreceptor genes of the waterflea *Daphnia pulex*: many Grs but no Ors. BMC Evolutionary Biology, 9, 79.

Pereira, H.S. & Sokolowski, M.B. (1993). Mutations in the larval foraging gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America, 90, 5044-5046.

Peres-Neto, P.R., Jackson, D.A. & Somers, K.M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. Comput. Stat. Data Anal., 49, 974-997.

Persson, J., Brett, M.T., Vrede, T. & Ravet, J.L. (2007). Food quantity and quality regulation of trophic transfer between primary producers and a keystone grazer (*Daphnia*) in pelagic freshwater food webs. Oikos, 116, 1152-1163.

Peters, R.H. (1986). The ecological implications of body size. Cambridge University Press.

Pfrender, M.E. & Lynch, M. (2000). Quantitative Genetic Variation in *Daphnia*: Temporal Changes in Genetic Architecture. Evolution, 54, 1502-1509.

Pfrender, M.E. (2012). Triangulating the genetic basis of adaptation to multifarious selection. Mol. Ecol., 21, 2051-2053.

Pijanowska, J., Weider, L.J. & Lampert, W. (1993). Predator-mediated genotypic shifts in a prey population: experimental evidence. Oecologia, 96, 40-42.

Porter, K.G., Gerritsen, J. & Orcutt, J.D. (1982). The Effect of Food Concentration on Swimming Patterns, Feeding Behavior, Ingestion, Assimilation, and Respiration by *Daphnia*. Limn. Ocean., 27, 935-949.

R Development Core Team. (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reede, T. (1995). Life history shifts in response to different levels of fish kairomones in *Daphnia*. *J. Plank. Res*., 17, 1661-1667.

Renn, S.C.P., Aubin-Horth, N. & Hofmann, H.A. (2008). Fish and chips: functional genomics of social plasticity in an African cichlid fish. J Exp Biol, 211, 3041-3056.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., et al. (2011). Detecting Novel Associations in Large Data Sets. Science, 334, 1518 -1524.

Reznick, D., Nunney, L. & Tessier, A. (2000). Big houses, big cars, superfleas and the costs of reproduction. Trends Ecol. Evol., 15, 421-425.

Rhee, S.Y., Wood, V., Dolinski, K. & Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nat Rev Genet*, 9, 509-515.

Robinson, G.E., Fernald, R.D. & Clayton, D.F. (2008). Genes and social behavior. science, 322, 896-900.

Rockman, M.V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. Nature, 456, 738-744.

Roff, D.A. & Fairbairn, D.J. (2007). The evolution of trade-offs: where are we? J. Evol. Biol., 20, 433-447.

Roff, D.A. (1992). The evolution of life histories: theory and analysis. Springer.

Roff, D.A. (2007). Contributions of genomics to life-history theory. Nat Rev Genet, 8, 116-25.

Rosenblum, E.B., Römpler, H., Schöneberg, T. & Hoekstra, H.E. (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proc. Natl. Acad. Sci., 107, 2113-2117.

Rosenheim, J.A., Heimpel, G.E. & Mangel, M. (2000). Egg maturation, egg resorption and the costliness of transient egg limitation in insects. Proc. R. Soc. Lond. B Biol. Sci., 267, 1565-1573.

Rouvalis, A., Karadima, C., Zioris, I., Sakkas, V., Albanis, T. & Iliopoulou-Georgudaki, J. (2009). Determination of pesticides and toxic potency of rainwater samples in western Greece. Ecotoxicology And Environmental Safety, 72, 828-833.

Sarnelle, O. (1993). Herbivore effects on phytoplankton succession in a eutrophic lake. Ecol. Monogr., 129-149.

Sarnelle, O. (2005). *Daphnia* as keystone predators: effects on phytoplankton diversity and grazing resistance. J. Plankton Res., 27, 1229.

Scheiner, S.M. & Berrigan, D. (1998). The genetics of phenotypic plasticity. VIII. The cost of plasticity in *Daphnia pulex*. Evolution, 368-378.

Scoville, A.G. & Pfrender, M.E. (2010). Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. Proc. Natl. Acad. Sci., 107, 4260 -4263.

Seger, R. & Krebs, E.G. (1995). The MAPK signaling cascade. Faseb J., 9, 726-735.

Seuront, L., Brewer, M.C., Strickler, J.R. & others. (2004). Quantifying zooplankton swimming behavior: the question of scale. Handb. Scaling Methods Aquat. Ecol. Anal. Simul. CRC Press Boca Raton, 333-359.

Seuront, L., Brewer, M.C., Strickler, J.R. & others. (2004). Quantifying zooplankton swimming behavior: the question of scale. Handbook of scaling methods in aquatic ecology: measurements, analysis, simulation. CRC Press, Boca Raton, 333-359.

Shimizu, Y. & Urabe, J. (2007). Regulation of phosphorus stoichiometry and growth rate of consumers: theoretical and experimental analyses with *Daphnia*. Oecologia, 155, 21-31.

Simon, J.-C., Pfrender, M.E., Tollrian, R., Tagu, D. & Colbourne, J.K. (2011). Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. J. Hered., 102, 512 -525.

Simon, J.-C., Pfrender, M.E., Tollrian, R., Tagu, D. & Colbourne, J.K. (2011). Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. Journal of Heredity, 102, 512 -525.

Skunca, N., Altenhoff, A. & Dessimoz, C. (2012). Quality of Computationally Inferred Gene Ontology Annotations. Plos Comput Biol, 8, e1002533.

Smith, C.C. & Fretwell, S.D. (1974). The Optimal Balance between Size and Number of Offspring. Am. Nat., 108, 499-506.

Spitze, K. (1991). *Chaoborus* Predation and Life-History Evolution in *Daphnia pulex*: Temporal Pattern of Population Diversity, Fitness, and Mean Life History. Evolution, 45, 82-92.

Spitze, K. (1992). Predator-Mediated Plasticity Of Prey Life-History And Morphology - *Chaoborus*-Americanus Predation On *Daphnia-Pulex*. Am. Nat., 139, 229-247.

Spitze, K. (1995). Quantitative genetics of zooplankton life histories. Experientia, 51, 454-464.

Spitze, K., Burnson, J. & Lynch, M. (1991). The covariance structure of life-history characters in *Daphnia pulex*. Evolution, 45, 1081-1090.

Stearns, S.C. & Magwene, P. (2003). The Naturalist in a World of Genomics. Am. Nat., 161, 171-180.

Stern, D. & Orgogozo, V. (2009). Is genetic evolution predictable? SCIENCE, 323, 746.

Stern, D.L. & Orgogozo, V. (2008). The Loci of Evolution: How Predictable is Genetic Evolution? Evolution, 62, 2155-2177.

Sterner, R.W. & Elser, J.J. (2002). Ecological stoichiometry: the biology of elements from molecules to the biosphere. Princeton University Press.

Stibor, H. (2002). The role of yolk protein dynamics and predator kairomones for the life history of *Daphnia magna*. Ecology, 83, 362-369.

Stocker, H. & Hafen, E. (2000). Genetic control of cell size. Curr. Opin. Genet. Dev., 10, 529-535.

Stone, E.A. & Ayroles, J.F. (2009). Modulated Modularity Clustering as an Exploratory Tool for Functional Genomic Inference. PLoS Genet., 5, e1000479.

Stuart, J.M., Segal, E., Koller, D. & Kim, S.K. (2003). A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. Science, 302, 249-255.

Stuefer, J.F., Van Hulzen, J. b. & During, H.J. (2002). A genotypic trade-off between the number and size of clonal offspring in the stoloniferous herb *Potentilla reptans*. J. Evol. Biol., 15, 880-884.

Supek, F., Bo njak, M., Skunca, N. & Smuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE, 6, e21800.

Székely, G.J. & Rizzo, M.L. (2009). Brownian distance covariance. Ann. Appl. Stat., 1236-1265.

Székely, G.J. & Rizzo, M.L. (2013). The distance correlation -test of independence in high dimension. J. Multivar. Anal., 117, 193-213.

Székely, G.J., Rizzo, M.L. & Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. Ann. Stat., 35, 2769-2794.

Tanaka, E., Bailey, T., Grant, C.E., Noble, W.S. & Keich, U. (2011). Improved similarity scores for comparing motifs. Bioinformatics, 27, 1603-1609.

Taub, M.A., Corrada Bravo, H. & Irizarry, R.A. (2010). Overcoming bias and systematic errors in next generation sequencing data. Genome Med, 2, 87.

Tessier, A. & Goulden, C. (1982). Estimating food limitation in cladoceran populations. Limnol. Ocean., 27, 707-717.

Tessier, A., Young, A. & Leibold, M. (1992). Population dynamics and body-size selection in *Daphnia*. Limnol. Ocean., 37, 1-13.

Tessier, A.J. & Consolatti, N.L. (1989). Variation in Offspring Size in *Daphnia* and Consequences for Individual Fitness. Oikos, 56, 269-276.

Tessier, A.J. & Consolatti, N.L. (1991). Resource Quantity and Offspring Quality in *Daphnia*. Ecology, 72, 468-478.

Tessier, A.J. & Leibold, M.A. (1997). Habitat use and ecological specialization within lake *Daphnia* populations. Oecologia, 109, 561-570.

Thompson, J.N. (1998). Rapid evolution as an ecological process. Trends Ecol. Evol., 13, 329-332.

Tilden, A.R., McCoole, M.D., Harmon, S.M., Baer, K.N. & Christie, A.E. (2011). Genomic identification of a putative circadian system in the cladoceran crustacean *Daphnia pulex*. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 6, 282-309.

Tilman, D. (1982). Resource competition and community structure. Princeton University Press.

Tollrian, R. & Leese, F. (2010). Ecological genomics: steps towards unraveling the genetic basis of inducible defenses in *Daphnia*. Bmc Biol., 8, 51.

Tollrian, R. (1993). Neckteeth formation in *Daphnia pulex* as an example of continuous phenotypic plasticity: morphological effects of *Chaoborus* kairomone concentration and their quantification. J. Plankton Res., 15, 1309-1318.

Trivers, R.L. (1974). Parent-offspring conflict. Am. Zool., 14, 249-264.

Turner, T.L., Stewart, A.D., Fields, A.T., Rice, W.R. & Tarone, A.M. (2011). Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. Plos Genet., 7, e1001336.

Urabe, J., Kyle, M., *Magna*, W., Yoshida, T., Andersen, T. & Elser, J. (2002). Reduced light increases herbivore production due to stoichiometric effects of light/nutrient balance. Ecology, 83, 619-627.

Van Gool, E. & Ringelberg, J. (1998). Light-induced migration behaviour of *Daphnia* modified by food and predator kairomones. Animal behaviour, 56, 741-747.

Wagner, A. (1996). Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biol. Cybern., 74, 557-567.

Wang, S., Meyer, E., McKay, J.K. & Matz, M.V. (2012). 2b-rad: a simple and flexible method for genome-wide genotyping. Nat. Methods, 9, 808-810.

Ware, D.M. (1973). Risk of epibenthic prey to predation by rainbow trout (Salmo gairdneri). Journal of the Fisheries Board of Canada, 30, 787-797.

Webb, C.T., Hoeting, J.A., Ames, G.M., Pyne, M.I. & LeRoy Poff, N. (2010). A structured and dynamic framework to advance traits-based theory and prediction in ecology. Ecol. Lett., 13, 267-283.

Weber, A. & Van Noordwijk, A. (2002). Swimming behaviour of *Daphnia* clones: differentiation through predator infochemicals. *J. Plank. Res*., 24, 1335 -1348.

Weider, L.J., *Magna*, W., Acharya, K., Glenn, K.L., Kyle, M., Urabe, J., et al. (2005). Genotype X environment interactions, stoichiometric food quality effects, and clonal coexistence in *Daphnia pulex*. Oecologia, 143, 537-547.

Weismann, A. (1876). Beiträge zur Naturgeschichte der Daphnoiden. W. Engelmann.

West, G.B., Brown, J.H. & Enquist, B.J. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organisms. Science, 284, 1677-1679.

Wetterer, J.K. (1989). Mechanisms of prey choice by planktivorous fish: perceptual constraints and rules of thumb. Anim. Behav., 37, 955-967.

Wetzel, R.G. (2001). Limnology: lake and river ecosystems. Academic press.

Wheat, C. (2010). Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica, 138, 433-451.

Winder, M., Spaak, P. & Mooij, W.M. (2004). Trade-offs in *Daphnia* habitat selection. Ecology, 85, 2027-2036.

Wolinska, J., Löffler, A. & Spaak, P. (2007). Taxon-specific reaction norms to predator cues in a hybrid *Daphnia* complex. Freshw. Biol., 52, 1198-1209.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., et al. (2003). The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol., 20, 1377-1419.

Wright, D.I. & O,ÄôBrien, W.J. (1982). Differential location of *Chaoborus* larvae and *Daphnia* by fish: the importance of motion and visible size. Am. Midl. Nat., 68-73.

Yarmolinsky, D.A., Zuker, C.S. & Ryba, N.J.P. (2009). Common Sense about Taste: From Mammals to Insects. Cell, 139, 234-244.

Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F. & Hairston, N.G. (2003). Rapid evolution drives ecological dynamics in a predator-prey system. Nature, 424, 303-306.